GA1.13:PEMD-88-27

United States General Accounting Office

GAO

Report to the Chairman, Committee on Post Office and Civil Service, House of Representatives

[illegible]
93

August 1988

FEDERAL WORKFORCE

A Framework for Studying Its Quality Over Time

GAO

United States
General Accounting Office
Washington, D.C. 20548

Program Evaluation and
Methodology Division

B-228638

August 4, 1988

The Honorable William D. Ford
Chairman, Committee on Post Office and Civil Service
House of Representatives

Dear Mr. Chairman:

At your request, we have examined the feasibility of assessing the quality of the federal civilian workforce. As was agreed, we focused on professional and administrative staff. Our conclusion is that such an assessment can be made.

In order to reach this conclusion, we developed a measurable definition of quality centered on attributes of the individual and the match of the individual's capabilities to the needs of the job. We examined agency personnel data to see whether indicators of quality pertinent to this definition were readily available and found that they were not. We then developed proposals for new data-gathering that would permit policymakers and others to obtain answers to specific common questions about the quality of those entering and leaving the professional and administrative workforce, and about changes in overall quality across time.

As arranged with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution of it until 30 days from the date of the report. At that time we will send copies to the Office of Personnel Management and others who are interested and make copies available to others upon request.

Sincerely yours,

Eleanor Chelimsky
Director

Executive Summary

Purpose

The federal workforce costs taxpayers billions of dollars annually and affects all of our lives through the performance of its functions. A generation or more ago, so beliefs run, public service was a high calling, attracting the best and the brightest. While such a golden age perhaps never existed, it is clear that in recent years concern has intensified about changes in the attractiveness of public service and possible consequences for the quality of the federal workforce.

The House Committee on Post Office and Civil Service asked GAO to (1) to examine the feasibility of producing generalizable and reliable information on the quality of the federal workforce over time and (2) in the event this was feasible, to outline a design for obtaining it.

Background

The overall federal civilian workforce (including legislative, judicial, and executive branch employees and employees of the U.S. Postal Service) includes over 3 million people nationwide, about 3 percent of all civilian employees in the United States. Of the non-postal workforce of 1.85 million, 42 percent is concentrated in professional and administrative positions such as accountant, attorney, engineer, or contract specialist. GAO's review addresses the executive branch civilian workforce in professional and administrative positions.

The quality of the federal workforce is important because it presumably affects performance: that is, government agencies' ability to carry out their responsibilities effectively and economically. Direct measures of the performance of some government functions are possible, as shown in productivity data reported by the Bureau of Labor Statistics. Trends in such data, however, can reflect a number of different influences including increased automation, improved telecommunications, or management reforms, along with changes in the workforce. Thus, data on performance could not address the issue of concern to the committee: the effectiveness of government efforts to recruit, develop, and retain a workforce with the quality needed for efficient performance of government functions.

Results in Brief

For the purpose of selecting indicators, GAO developed a two-part definition of quality including both employee capabilities and the degree to which those capabilities are matched to the requirements of a particular job. Judging whether employees in a particular occupation are of high quality thus involves not only n absolute level of knowledge, skill, and ability, but also congruence between those capabilities and the needs of the work. (See chapter 2.)

GAO found that the information currently available in agency personnel records was not comprehensive, accessible, or current enough for the purposes of the quality assessment effort. GAO also found that it would not be practical to change agency systems so that they might provide a better data base for the purpose in the future. (See chapter 3.)

GAO recommends an alternative approach, to be implemented by the executive branch, based on new data to be gathered chiefly by surveying samples of employees in selected occupations. The proposed data would permit generalized answers to many of the major questions that have been raised about quality in the three different segments of the workforce. (See chapters 4-6.) The approach appears to be relatively low in direct federal costs (see appendix III), and it has the potential to be refined and expanded over time as experience accumulates on its strengths and limitations. (See chapter 7.)

GAO's Analysis

GAO's definition of quality includes two basic concepts: the capabilities of the individual and the match between the extent of those capabilities and the needs of the specific job. To define and measure an individual's capabilities, GAO selected two basic characteristics: knowledge, skill and ability; and attitudes, values, and motivation. There are many types of data which can be used to show each of these, including education, training, test scores and licenses as indicators of knowledge, skill, and ability, and individuals' views as indicators of attitudes, values, and motivation.

GAO believes an indicator of the match between individuals' capabilities and their current work should be obtained, at least initially, from ratings by both individual employees and their supervisors. (See chapter 2.)

Existing Personnel Data

After reviewing the data maintained on employees in all forms at all levels in eight agencies, GAO concluded that only a few aspects of its definition of quality could be measured from existing records. Automated files contain a few indicators pertinent to workforce quality, chiefly on education. These are, however, not up to date. Apart from scattered surveys of small groups, no data exist that could shed light on other parts of the definition of quality such as attitudes, values and motivation or the match of capabilities to the needs of the job. Some relevant information might be found in official personnel files, but would be costly to retrieve. There appear to be significant barriers to expanding agency personnel data systems to add routine collection and storage of additional data on GAO's definition of quality. The burden of data-collection and data-entry would be too great, and the motivation within the agencies to achieve completeness and accuracy too low, to yield worthwhile information. (See chapter 3.)

Sample Studies of Entering, Current and Exiting Employees

The major questions that an adequate assessment of workforce quality must answer are: How good are those being attracted to government jobs and how do they compare to those taking jobs elsewhere? What changes have there been in the quality of the workforce over time? Are those leaving government different in quality from those who do not leave? GAO's design proposes that new data on indicators of quality be gathered from samples of employees using surveys. This approach would for the first time provide regular, comparable information addressing each of the three questions for major occupations in the professional and administrative workforce. The costs that can be estimated appear to be modest, less than $1 million for an initial cycle of data-gathering and analysis. (See appendix III.)

The assessment of the newly-hired part of the workforce poses special challenges, since proper analysis requires a comparison group of non-federal employees in the same occupations. GAO presents several options for methodologies of obtaining such a sample and the costs are uncertain since nothing similar has been done before. The data would help answer such questions as how people the federal government hired in specific occupations compared, in terms of educational achievement or experience, to those entering similar jobs in non-federal sectors. (See chapter 4.)

Evaluation of change over time in the quality of the current workforce is the central element of the GAO assessment design, involving surveying a sample of current employees at three levels of experience in 23 occupations. The data would answer such questions as whether, over time, federal workers in specific jobs show changes in the extent or quality of formal education they bring to their work, the extent or quality of continuing education and training, whether over time those in an occupational series are generally more or less experienced, and to what extent the workers and their supervisors believe that individual workers' capabilities are adequate to their jobs. (See chapter 5.)

The third question, about the quality of those separating from federal jobs, requires a comparison with employees of similar experience levels remaining in similar jobs. Data from this part of the assessment would answer questions about whether federal occupations were losing individuals with higher qualifications, faster track records, and greater past recognition than those retained. (See chapter 6.)

GAO suggests that a lead agency be selected on criteria of technical expertise, access to the data bases needed, independence, commitment to the effort, and availability of resources. The core assessment covering a consistent set of occupations should be repeated at least every four years to provide trend data on changes in quality. (See chapter 7.)

GAO concludes that methodologically adequate answers to questions about the quality of the federal workforce and changes in it over time will not come from existing personnel records or existing productivity studies. Useful data could, however, be provided by a new program of data-gathering and analysis, established at congressional direction and with sufficient resources to carry out the work in a methodologically sound way.

Recommendation

GAO recommends that the Congress authorize a continuing program to assess workforce quality based on the framework outlined in this report.

Agency Comments

GAO discussed the proposed assessment with officials at the Office of Personnel Management and the Merit Systems Protection Board. The officials in general raised no major problems with the proposed definition of quality, the initial indicators, or the broad feasibility of the designs for the three segments of the assessment. However, in accordance with the requester's wishes, we did not obtain written comments on a draft of this report.

Contents

Executive Summary 2

Chapter 1 Introduction 10
Workforce Quality: Glimpses of the Information Problem 11
Objective, Scope, and Methodology 13
Organization of the Report 15

Chapter 2 The Definition of Quality 16
Possible Definitions 16
Assessing Quality in Individuals in the Workforce 22
Definition of Workforce Quality 25
Specific Indicators of Quality 26
Summary and Conclusions 29

Chapter 3 Using Administrative Data to Assess Workforce Quality 31
Data That Are Available in Agency Record Systems 31
Feasibility of Enlarging Administrative Data 34
Data That Are Available in OPM Record Systems 36
Modifications in OPM Data 37
Summary and Conclusions 38

Chapter 4 Assessing the Entering Workforce 39
Questions About the Entering Workforce 39
Data Needed to Answer the Questions and Potential Sources 40
Proposed Design 43
Summary and Conclusions 48

Chapter 5 Assessing Changes in the Workforce 50
Questions About the General Workforce 50
Data Needed to Answer the Questions and Potential Sources 51
Proposed Design 53
Summary and Conclusions 57

Chapter 6 Assessing Separations From the Workforce

58

Questions About Those Leaving the Workforce 58

Data Needed to Answer the Questions and Potential Sources 59

Proposed Design 62

Summary and Conclusions 64

Chapter 7 Strategies for Implementing the Assessment

66

Organizational Arrangements 67

Funding for the Assessment 70

Timing and Phasing of the Assessment 71

Conclusions 72

Recommendation to the Congress 73

Agency Comments 73

Appendixes

Appendix I: Congressional Request Letter 74

Appendix II: Selection of Occupations for Sampling 76

Appendix III: Sample Sizes and Costs 79

Appendix IV: Advisory Panel Members, and Others Consulted During Our Work 86

Tables

Table 2.1: Our Evaluation of Types of Measures for Three Dimensions of Workforce Quality 28

Table 2.2: Potential Indicators for Different Parts of the Definition of Workforce Quality 29

Table 3.1: Number of Agencies Responding Positively to Questions About Workforce Data Elements 32

Table 3.2: Number of Agencies Citing Barriers to Enlarging Automated Personnel Data Systems 35

Table 3.3: Number of Agencies Citing Specific Concerns About Enlarged Data Systems for Monitoring Quality 36

Table 4.1: The Feasibility and Interpretability of Alternative Data Sources for Assessing New Employees 41

Table 4.2: Dimensions, Indicators, and Data Sources for Assessing New Employees 47

Table 5.1: The Feasibility and Interpretability of Alternative Data Sources for Assessing the General Workforce 52

Table 5.2: Dimensions, Indicators, and Data Sources for Assessing the General Workforce 55

Table 6.1: The Feasibility and Interpretability of Alternative Data Sources for Assessing Separated and Retained Employees 61
Table 6.2: Dimensions, Indicators, and Data Sources for Assessing Separated and Retained Employees 64
Table 7.1: Design for an Assessment of Federal Workforce Quality 66
Table 7.2: Relative Difficulty of Parts of the Assessment 72
Table II.1: Occupations Selected for Quality Assessment 77
Table II.2: Additional Occupations to Be Grouped for Quality Assessment 78
Table III.1: Sample Sizes Needed for Three Segments of the Proposed Assessment of Federal Workforce Quality 81
Table III.2: Estimated Costs of the Proposed Assessment of Federal Workforce Quality 85

Figure

Figure 2.1: Workforce Quality and Its Larger Context 17

Abbreviations

BLS	Bureau of Labor Statistics
CPDF	Central Personnel Data File
EPA	Environmental Protection Agency
GAO	General Accounting Office
HHS	Department of Health and Human Services
MSPB	Merit Systems Protection Board
NASA	National Aeronautics and Space Administration
OPM	Office of Personnel Management
PACE	Professional and Administrative Career Examination
VA	Veterans Administration

BLANK PAGE

Introduction

There is a long history of concern about the quality of the federal workforce. The statutory merit principles require that "selection and advancement should be determined solely on the basis of relative ability, knowledge, and skills." The federal government has invested great effort and has provided leadership to both the public and private sectors in developing methods for fairly assessing individual qualities to support selection of staff based on merit alone. Beyond selection, the present or desired quality of the workforce also is discussed in the context of many other personnel policy issues including setting pay, classifying positions, evaluating performance, linking pay with performance, and defining the benefits that will be offered those who are retiring. Workforce quality is brought into the more general debates about the ability of government to perform its missions, especially as government increasingly competes for employees in shortage occupations.[1]

Despite the frequency of discussion of the subject, no statutory guidelines exist that direct comprehensive or regular assessment of the quality of the federal workforce. Rather, it is analyzed using individual anecdotes, unsystematic sampling of employee or supervisor opinions, or by inferences from other data such as comparisons of federal and nonfederal employees' pay and benefits. Such evidence has been used to make claims that the federal government is not attracting as strong candidates for jobs as it used to, that the overall quality of the workforce is declining, or that highly valued employees are leaving government in recent years at a rate that should be cause for alarm.

The Committee on Post Office and Civil Service of the House of Representatives concluded that discussion of many topics in federal personnel policy could benefit if better information could be obtained on the quality of the workforce, both its current status and changes over time. The committee request (see appendix I) asked us to examine the feasibility of setting up such a data base and, if it seemed feasible, to propose a design.

---

[1] The most recent listing of GAO work on federal personnel management is in U.S. General Accounting Office, Federal Personnel: Annual Report on Activities of OPM and MSPB, Fiscal Year 1986, GAO/GGD-87-57 (Washington, D.C.: May 1987). The Comptroller General testified on GAO work related to revitalizing public service (GAO/T-GGD-88-21, March 24, 1988).

Workforce Quality: Glimpses of the Information Problem

Scattered efforts to evaluate the quality of the workforce, although intriguing and provocative, also suggest the problems of attempting the task. We found the following examples of assessments of quality in groups entering federal jobs, in the workforce more generally, or in those leaving federal jobs:

The Internal Revenue Service studied the quality of their incoming revenue agents, compared to new staff of accounting firms, by giving a national test of accounting knowledge and found that 84 percent of the comparison group outside of government scored higher than the mean of the federal group in each of two years of testing.

The Merit Systems Protection Board (MSPB) examined the quality of the workforce by asking managers' views on whether the quality of applicants for professional and administrative jobs at entry-, mid-, and senior levels had changed in the four years preceding the 1986 survey, and found that 35-36 percent believed it had worsened at each level; 33-40 percent believed it had stayed the same; and 24-31 percent believed it had improved (the percentages varied slightly across the levels included in the question).

We attempted to profile the procurement workforce by using available automated personnel files and reported that there was information from a data base in the Office of Personnel Management (OPM) on employees' education at the time of hiring that showed an increase over time in the proportion of employees with college degrees. However, no data were available on current status or trends in other indicators believed to be essential, such as experience and training.

We reviewed the quality of technical staff assigned to the complex tasks of environmental cleanup in the Superfund section of the Environmental Protection Agency (EPA) by asking employees to give opinions of shortages of critical skills and found that 39 percent of employees had a problem assuring the quality of their work because of a lack of skilled personnel.

The Congressional Budget Office studied the quality of managers and supervisors who left government compared to those who stayed by examining their performance appraisal scores and found that because those with lower ratings left at a higher rate than average, perhaps turnover was leading to separation of the less satisfactory performers.

The Department of Defense studied the quality of science and engineering personnel who left the department, compared to those who remained, finding that separation rates of those who had scored above 650 on the mathematics part of the Scholastic Aptitude Test (at the time of graduating from high school) were about 50 percent higher than those of persons who scored below that level.

We assessed the effectiveness of a wide range of human resource management policies by surveying a government-wide sample of personnel officers, including one question about their opinions of changes in their ability to acquire and retain quality personnel in the last five years, and found that half believed their ability to acquire good people was worse or much worse and over 60 percent believed their ability to retain such people was worse or much worse.

We studied whether exceptional Marine Corps officers were being retained in the Corps by examining performance appraisals and reenlistment records and found that the service did not lose its best men, at least not during the period reviewed.

Although useful for other purposes no doubt, the studies from which we drew these examples highlight a number of difficulties in pulling together current information for judging the quality of the workforce. In addition to asking different questions, the studies show great diversity of approach. They have not been guided by any law or regulation, have not been performed at predetermined intervals to allow trend analysis, have no common unit of analysis (having examined single occupations or clusters, at different levels of aggregation within part of one agency, within a department, or across government), have had different comparison groups or none, and have not been consistent in the indicators included in the definition of quality and how they were measured. Further, questions can be raised about individual indicators used in the studies. For example, is very much learned about the quality of experienced employees by examining their college entrance test scores? Do general tests of accounting knowledge, even when developed by the major national professional organization, tell us much about quality in relation to the tasks of a federal revenue agent? Are managerial performance appraisals done in different ways in different agencies comparable? Do we gain useful information from managers' general opinions about such a broad question as the quality of applicants or new hires, without reference to any occupation or to any standard of needed quality? Though these studies may have served other purposes well, they demonstrate the problem of trying to understand workforce quality from the existing fragmented sets of data.

Our challenge was to imagine an alternative information base that would remedy the many shortcomings of measurement, design, and interpretation that could be pointed out in the information now available.

Objective, Scope, and Methodology

The objective of our effort, then, was to develop a means of routinely providing Congress and other parts of government with aggregated and analyzed descriptive data on the quality of the workforce over time.

The committee chairman's request referred to the civilian workforce in general terms. After agreement with committee staff, we confined our work to considering assessment of the executive branch civilian non-postal workforce. Thus we did not consider designs for assessing the quality of presidential appointees, the Senior Executive Service, the uniformed military, or legislative or judicial branch employees. We also agreed to focus only on professional and administrative occupational series within the civilian executive branch workforce. (The other two types of occupational series are technical and clerical.)

To carry out this objective we developed a definition of quality based on discussions with experts in measurement, federal employee groups, and with experienced personnel managers from federal and nonfederal sectors. (See appendix IV for a list of persons and organizations consulted.) We then surveyed eight federal agencies to discover the extent to which their civilian personnel records contained data that matched our definition of quality.[2] We asked the agencies also about any other data they had pertinent to alternative definitions of quality that we should consider. If few data were available, or were inaccessible, we asked about the feasibility of improvements in the agencies' overall data on their employees. We reviewed the data maintained in the OPM Central Personnel Data File (CPDF) and the feasibility of additions to that data base.

---

[2] The eight departments or agencies were: Air Force, Environmental Protection Agency, Health and Human Services (HHS), Interior, Justice, National Aeronautics and Space Administration (NASA), Navy, and Veterans Administration (VA). We chose agencies on two criteria: first, that they included substantial numbers of employees in a range of professional and administrative jobs, and second, that (in the cases of EPA, NASA and VA) they included occupations of special interest such as scientists, engineers, and health specialties.

To go beyond information now in agency records, we developed an alternative approach based on gathering new data from samples of federal employees and, where needed, comparison groups of other employees. To develop the proposal, we reviewed the public discussions of workforce quality and determined the main distinctive questions commonly addressed so that our design would allow the production of responsive information. For each question, we reviewed technical measurement alternatives including individual tests and measures used by OPM; existing surveys that could, with some augmentation, serve our purpose; OPM statistical data on the segments of the workforce needing to be assessed for each question; prior efforts to measure the various elements of our definition; statistical sampling requirements; and the costs and feasibility of our approach to each of the questions.

We selected an advisory group to review our work. The group included former high-level federal officials; federal agency personnel executives; an employee group official; experts in the disciplines of political science, evaluation, labor economics, and psychological measurement; a former member of Congress who was also formerly a state human resource executive; and an executive and personnel assessment expert from the private sector. We obtained advisory comments from these consultants on our overall approach to the project at the outset, and we reviewed the proposed assessment questions and designs with them as we completed our analysis. They also reviewed the draft report. Appendix IV lists the members of the group.

Our approach has several strengths, including the detail of our review of existing data stored in selected agencies' computer files and paper records and the breadth of our discussions with interested groups and individuals about issues in defining or measuring quality. One limitation is that we cannot suggest a design representing a rigorous consensus of experts' views, as we did not do a general national survey of opinions on the best way to approach the task. Another is that our design work addressed only the full-time permanent professional and administrative segment of the overall federal executive branch workforce. Although we believe it is one of the most important parts, and the appropriate group on which quality assessment should begin, it could be useful to devise relevant measurements and research designs to track the quality of those in appointed positions, in the Senior Executive Service, in clerical, technical, and wage-grade jobs, or with temporary or part-time appointments.

Organization of the Report

Chapter 2 reviews a range of possible areas of measurement that could be included in a definition of quality, gives the definition we propose, evaluates general types of data that might be used in measuring the elements of the definition, and concludes with specific indicators we looked for in agency data. Chapter 3 reports our review of personnel data maintained in agency records and our evaluation of its usefulness for measuring the proposed definition. Chapters 4-6 present three distinct questions about the workforce and designs that should provide useful data on each. Chapter 7 suggests some issues of implementation of the assessment. Appendices include: the congressional request letter (appendix I), occupations proposed to be sampled (appendix II), sample sizes and estimated costs for an initial cycle of the assessment (appendix III), and those we consulted (appendix IV).

Chapter 2

The Definition of Quality

The examples of scattered quality measurement efforts cited in chapter 1 show the variety of definitions and measures of quality that can be used, including individuals' knowledge (as demonstrated on a test) or job performance (as assessed by supervisors), or process measures such as difficulty of recruitment or retention (measured by managers' opinions). We found widespread belief that any assessment should start with a multi-faceted definition and that individual capacities should not be assessed in isolation from specific work tasks.

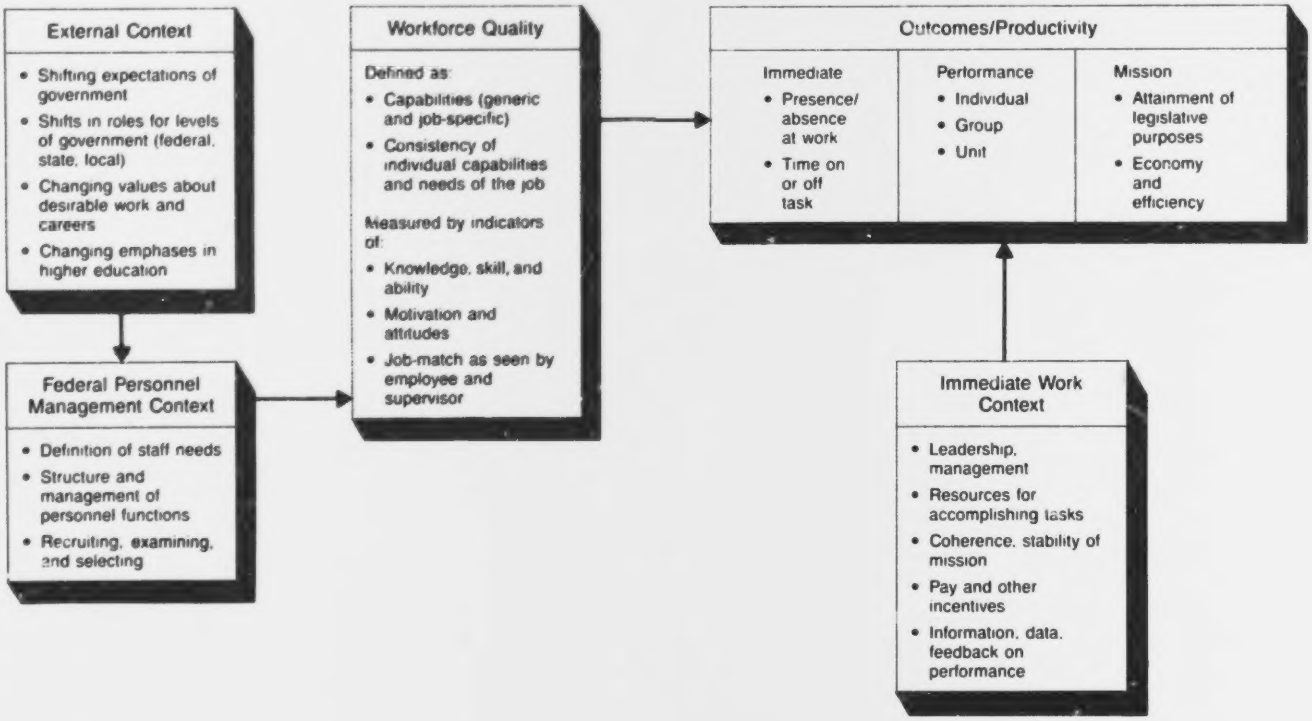
This chapter is organized into three sections. The next section examines the concept of a quality workforce and the wide range of information that has been used to address related questions and highlights the focus on individuals that guided the definition of quality proposed for this assessment. The second section analyzes various dimensions that could be included in the definition and gives the proposed two-part definition of a quality workforce as the possession of strong capabilities (general and job-specific) and the degree to which those capabilities are consistent with the needs of the specific job. The last section evaluates types of data that could be used to measure aspects of the definition and presents a list of specific data elements for which we searched in existing personnel data.

Possible Definitions

It is at first tempting to try to include in the definition of quality the wide range of continuing concerns about government: for example, its overall efficiency and productivity, the integrity of its officials, or the innovativeness of its scientific and engineering endeavor. The more the definition of quality is broadened, however, the more difficulty there is in measurement and interpretation. Restricting the assessment to individual characteristics provides a complex challenge in itself. Nevertheless, we began by considering a wide range of ideas that could be part of the definition of the quality of the federal workforce.

The federal workforce at any point in time is both a consequence of several environmental influences and a cause, along with other factors, of the results of government. We developed a model of the workforce, shown in figure 2.1, and examined each of the five elements of it for relevance to our assignment and for feasibility of measurement and interpretation. The shaded section of the figure, labelled workforce quality, including attributes of workers as individuals, became the focus of our measurement effort. The four sections that follow describe aspects of quality found in the four areas of the model that we eventually eliminated from consideration for inclusion in the definition: the

Figure 2.1: Workforce Quality and Its Larger Context[a]


External Context
Shifting expectations of government
Shifts in roles for levels of government (federal, state, local)
Changing values about desirable work and careers
Changing emphases in higher education
Federal Personnel Management Context
Definition of staff needs
Structure and management of personnel functions
Recruiting, examining, and selecting
Workforce Quality
Defined as:
Capabilities (generic and job-specific)
Consistency of individual capabilities and needs of the job
Measured by indicators of:
Knowledge, skill, and ability
Motivation and attitudes
Job-match as seen by employee and supervisor
Outcomes/Productivity
Immediate
Presence/absence at work
Time on or off task
Performance
Individual
Group
Unit
Mission
Attainment of legislative purposes
Economy and efficiency
Immediate Work Context
Leadership, management
Resources for accomplishing tasks
Coherence, stability of mission
Pay and other incentives
Information, data, feedback on performance

[a] Proposed indicators of workforce quality discussed later in this chapter are drawn from the area highlighted by shading

external context, the personnel management context, the immediate work context, and the outcomes of work.

The External Context

Several aspects of the broad environment in which the federal government competes with other employers for workers have been suggested as areas for improved measurement in order to understand workforce quality. For example, there are periodic shifts in public views about the role of government and about work and careers. As government responds to these changing expectations, there will be changes in the kinds of jobs that are available. In addition, there are shifts in the kinds of rewards people want from jobs. Both will influence job-seekers' views of the attractiveness of government work, and, hence, the nature of the applicant pool. Constantly rising costs of education are one specific influence affecting graduates' choices of jobs in recent years.[1]

The opinions of students, higher education faculty, or the public about different employers and different kinds of work can be readily studied. However, the importance of these views, among other influences on actual job choices, is unknown, and the data offer no direct answer to questions about workforce quality. Once particular quality problems have been identified, such as in particular occupations or career stages, then targeted studies of the external context can be designed to search for possible causes, if desired.

The Personnel Management Context

Common substitute measures for workforce quality are the processes of workforce planning, and the steps of recruiting, selecting, and classifying workers. In this view, one has only to judge the degree of independence, fairness, and meritocracy built into personnel management to have a yardstick for the quality of the resulting workforce.

It would be possible to evaluate the planning stage, examining the precision and validity of agencies' definitions of staff needs and the extent to which any definition of quality was used in workforce planning. In later stages of personnel management, managers and personnel staff advertise, recruit, interview, examine, and select individuals to meet those needs. Evaluation again could examine procedural details of each of these steps that affect the workforce actually hired and set to work. Feasibility, then, is not the problem here. Rather the difficulty is that we cannot be sure of the degree to which the quality of personnel management does in fact determine the quality of the workforce. That is, as with data on the external context, such information can only answer questions about workforce quality indirectly. Yet the need here is for more direct indicators about the workforce, to facilitate agreement about results.

---

[1] The external context that affects workforce quality can be conceived even more broadly to include the biological world we inhabit. One search for explanations for a decline in the quality of military recruits related that to a wide range of other health and environmental degradations ranging from maternal and pre-natal nutrition and changes in methods of delivering babies, to atomic fallout, diet, and food additives. The authors of this analysis concluded that this ecological context was the proper level of analysis and that "the quality decline...is an expectable and perhaps inevitable consequence of an accumulation of ecological changes that have occurred throughout the United States in the past several decades." The level of context at which one seeks causal factors clearly affects the search for remedies as well; these authors argued that the only proper response to the quality problem was to focus on biological research and interventions centered on diet and nutrition. (Bernard Rimland and Gerald Larson, "The Manpower Quality Decline: An Ecological Perspective," Armed Forces and Society, 8 (Fall 1981), 21-78.)

Independent sources of indicators are needed for a second reason, to provide an independent way to evaluate the processes themselves and especially variations in their quality. The most convincing causal evaluations would link measurable differences in the steps of personnel management to differences in work results (for example, to show that better selection tests result in a workforce better matched to tasks, which in turn is more productive than one less carefully selected). But in the absence of such strong study designs and the diverse data they would require, at least one would like to know, for example, whether increased effort in recruitment or more careful screening was associated with any noticeable change — that is, was there any contemporaneous change in any index of workforce knowledge, skill, or any other quality indicators? Thus it seemed important to develop an assessment design that has measures independent of personnel management, first, to provide a long-term set of indicators that can reveal broad problems to which management could respond, and second, to determine whether quality indicators show any trends that can be associated with government-wide management changes.

The Immediate Work Context

As in any workforce, federal employees are given specific work to do; they are led and supervised, offered short- and long-term rewards for their efforts, given resources to attain governmental objectives, and provided some degree of stability and support in their work environment. Each of these dimensions is variable and each can serve as an explanation for the quality of workers attracted and retained and for the quality of their work output.

For example, employee groups emphasized in interviews the importance of management in designing work and creating the conditions for effective or ineffective use of a workforce of whatever skill level. (Classification of jobs is said to be difficult to change quickly enough to keep pace with changing work.) Federal managers emphasize the fluctuating resources they are given and the shifts in agency policy that affect their efforts to deliver. More generally, there is wide discussion of the effects of newly-revised federal retirement benefits on who will be attracted and retained, of federal pay and its comparability to other sectors, and of other rewards such as promotions and merit increases. Tasks and results can be affected by the degree of thoughtful management evaluation to correct course based on past experience. All of these aspects of work setting and management may influence the quality of employees attracted and retained by the federal government and their work performance. Such analyses are not inappropriate; indeed, they are crucial in understanding how the government makes use of all of its resources, including its employees. But again, data on all these details of the context of management and leadership do not provide indicators pertinent to the committee's specific interest in assessing the workforce.

Outcomes/Productivity

The workforce is set to the wide range of government tasks, for example collecting taxes, writing checks, exploring space, gathering statistics, delivering mail, forecasting the weather, and keeping airplanes separated in the sky. It is appropriate to assess the quality of performance or results in each case.[2] We have called repeatedly for measurement of outcomes as an important source of data to improve government management, and arguably there could be no more important measure of the quality of the federal workforce than its performance.[3]

We decided not to propose measures of performance or outcomes in our response to the committee's request for two reasons. First, such data are difficult to gather even in one occupational area, and the assessment design needed to cover a spectrum of professional and administrative work. Second, outcome data could not be readily interpreted to show the quality of the human resources available for the government's tasks, which is the committee's interest. That is, it would be incorrect to draw conclusions about the workforce from data on work results, since those are affected by many factors in the situation (as outlined in figure 2.1) in addition to the knowledge, skill, ability, and other characteristics of the workforce.

Also, as in the case of personnel processes, accurate and comprehensive information on the workforce is important to collect in order to use it in analyzing variations in productivity data gathered separately. When problems with the outcomes of government are noted, where should we look to find what should be done? It would be useful to know whether changes had occurred in workforce quality, so that one could assess staffing as an area for added effort if outcome measures showed needs for improvement. That is, if there are smaller returns in tax audits, too many mid-air near misses, or increased patient complaints about hospital nursing care in government facilities, can these be understood better with reference to changes in the characteristics of the revenue agents, controllers, or nurses, or with reference to other factors (such as their assignment patterns and overwork, quality of supervision, effectiveness of technology, and so on)?

---

[2] As shown in figure 2.1, some intermediate outcomes include whether employees show up regularly, in condition to perform the assigned work, and stay with the job over a period sufficient to recoup the costs of getting them on board. As they attempt to produce results, employees spend time on or off task once they get on the job. And when people get to work, they produce results by themselves, in groups, and in larger units. These intermediate outcomes could all be evaluated through objective data. However, most people probably have direct services in mind as outcomes, or more generally, attainment of government missions.

[3] See, for example: U.S. General Accounting Office, Ways to Improve Federal Management and Use of Productivity Based Reward Systems, GAO/FPCD-81-24 (Washington, D.C.: Dec. 31, 1980) and Increased Use of Productivity Management Can Help Control Government Costs, GAO/AFMD-84-11 (Washington, D.C.: Nov. 11, 1983). Direct measures of the performance of some government functions are possible, as shown in productivity data reported by the Bureau of Labor Statistics.

We are proposing, in the definition outlined in the rest of this chapter and in the specific quality assessment efforts discussed in chapters 4-6, to focus on numerous attributes of the workforce. It is important to recognize, however, that their precise contributions to performance and outcomes are not settled by empirical research results, though they are plausible and widely discussed as important contributors. That is, it is still only an assumption — although a highly reasonable one — that efforts to increase the level of those attributes (such as by increasing pay or benefits in hopes of attracting or retaining workers with desired characteristics) will lead to improvements in work performance. Since there does exist some uncertainty over the importance of human resource variables for performance, it would be very useful to link the workforce quality assessment we are proposing to a continuing broader program of evaluation of performance, as well as research on its various causes.

We decided, however, to design the workforce quality assessment separate from such an overarching program of research on productivity because of cost, feasibility, and measurement uncertainties. Some people did recommend to us that we try to link descriptive data on the building blocks shown in figure 2.1 on individuals and their contexts together with measures of work performance and outcomes, to produce a causal analysis that could explain the contribution of each to the outcomes observed. Such a process of analysis is of course done in less formal, nonquantitative ways by a wide variety of actors and observers who weigh many factors to reach conclusions about how to use resources to accomplish government goals. But tidying all the many factors into an empirical model seemed extremely difficult, given the current state-of-the-art in measurement for most parts of the model, and non-quantitative models are not easy to use for recognizing changes over time. This is not to say that such an analysis should not be planned for the future, given continued improvement in measures of both workforce attributes and other inputs as well as improved outcome measures. For the present, however, we conclude that such an effort is probably not feasible with current levels of scientific measurement and analysis.[1]

The Individual as the Focus of Workforce Quality

In sum, although there are important questions in each of these four areas of context and outcomes, the most appropriate focus for initial measurement still appears to be the individual employee. Even though the characteristics of the individual worker form only one building block among the many needed for complete understanding, no other domain appears promising to measure in response to the questions of likely interest to the committee. Thus we concluded that our design for assessment should center on an initial set of employee variables alone. If that assessment proves workable and useful, it will both stimulate, and form a key starting point for, wider analyses of the causes and consequences of workforce quality.

Assessing Quality in Individuals in the Workforce

An enormous variety of individual characteristics of workers could be assessed. Considering the specific focus on the workforce as individuals, we examined various frameworks to develop a detailed definition.

Diverse Concepts of Individual Quality

In the broader literature many conceptualizations of human qualities exist, including those of:

- the economist, who makes comparative evaluations of the "human capital" in different labor forces (such as in different parts of the world) with relatively crude numbers representing schooling and work experience;
- the psychologist, trained to judge individual mental ability, personality, motivation, or overall mental health with written, face-to-face, and situational measures;

---

[1] The field of personnel psychology has contributed many studies that are steps in the same direction, examining the validity of various measures of people's characteristics in predicting later events at work. The data against which the predictions are tested, however, are often only indirect measures such as tenure and turnover (how long a worker stayed on the job) or opinions (supervisors' ratings) rather than work results. Advances beyond such work require more concrete measures of work results and measures of a more complete set of causal factors (in addition to employee attributes) to explain the results.

- the business analyst, trying the new specialization of human resource accounting and attempting to value a firm's human assets for a balance sheet; or
- the journalist, drawing on interviews, observation, history, philosophy, and current opinion polls in writing for the general public about the character of a public servant.

In the specific federal context, the first of the statutory merit principles included in the Civil Service Reform Act of 1978 states that "selection and advancement should be determined solely on the basis of relative ability, knowledge, and skills," which suggests some descriptors of quality that might be measured.

The two major criteria we used in examining potential constructs and measures for our definition were

- the feasibility of measurement in the federal setting (whether such data could be readily obtained on sizable numbers of people at dispersed locations) and
- the interpretability of the data (whether relatively clear inferences could be drawn from data on an indicator).

Further, we examined whether our definition should be influenced by the strict legal tests that apply to definitions and indicators of quality used in individual hiring and promotion decisions.

The Legal Standards of Definitions

The legal standards pertinent here reflect, on the one hand, continuation of the original aim of all merit selection methods by demanding closer scrutiny of whether tests or other screens fairly assess individuals' fitness for a particular job. The trend here is to require ever-stronger demonstrations that a specific quality dimension (for example, verbal ability), and the specific measurement of that dimension (in the same example, a paper-and-pencil reading comprehension or verbal analogies test), are related to performance of the tasks of a specific job.

On the other hand, other laws and court cases provide for review of any quality measurement if its effect, when used for selection, seems unfair to groups in the society, such as women or racial minorities. Thus, quality dimensions and measures chosen by employers with an eye to efficiently matching jobs and individuals face review on equity grounds as well as for evidence that they validly reflect the work.[5] The federal government discontinued the use of a test called the Professional and Administrative Career Examination (PACE) because of concern that minority groups scored disproportionately poorly on it.

The monitoring effort we were asked to design does not face the same rocky road as tests or other data used for employment selection. Because the information would have no impact on individuals, the validity of the assessment results would be tested not so much by the competing and conflicting legal standards of fairness just described, but by tests of practicality and usability by the consumers, who are policy makers in the legislative branch, leaders and managers of the workforce in the executive branch, and observers of government generally.

An example of the distinction between using indicators about individuals to predict something about them and using such data to monitor a larger system comes from the field of education. Concern about the dramatic downward slide in scores on the Scholastic Aptitude Test taken by high school students in the 1970s did not arise from, nor was it settled by, technical arguments over the test's precision in predicting college success (though that issue was of course raised). The aggregate figures were taken as a kind of national thermometer that was indicating something important about the health of education. Viewed that way, the data stimulated much further analysis of the nature of school achievement, its proper measure, and its multiple causes, and served as part of the impetus for a school reform movement that is still under way. In similar fashion, indicators of the quality of the workforce, even though not suitable as predictors about individuals, may nonetheless provide useful signals about an aggregate of employees and should provoke more detailed study to understand the message of initial warning flags.

Proper analysis and inference is a serious issue with respect to any proposed set of information. The issue of assigning the assessment work to an organization capable of credible interpretation of the assessment data is raised in chapter 7 in the discussion of implementation topics.

---

[5] The Committee on Ability Testing of the National Research Council analyzed the dilemmas of testing for abilities in education and employment in light of these two different legal standards. See Alexandra K. Wigdor and Wendell R. Garner (eds.), Ability Testing (Washington, D.C.: National Academy Press, 1982).

Definition of Workforce Quality

We ended our analysis of focus, feasibility, and measurement validity by centering the assessment on two broad domains of quality. We define a quality professional and administrative workforce as one with:

- strong capabilities (both generic and job-specific) and
- that degree of capability needed on a particular job.

These two complementary aspects of the definition of quality, using very different standards as their bases, emerged from our consultations with those who would either contribute data to the assessment or use the results. On one hand, quality can be defined as the extent to which an individual has desired characteristics such as education, job related experience, specific skills, motivation, or even certain personality traits. Gathering such data and arraying results by groups (such as how much education workers in a particular job have, or workers in a particular agency, or workers at a similar level of experience such as newcomers), carries the unstated assumption of a generic standard, that more of any characteristic would be desirable in the group. This approach to defining quality is exemplified by the question "Is the federal government attracting the best and the brightest?"

However, from a strategic point of view it is of equal or even greater interest to assess quality against a standard of need. That is, a manager thinking of particular tasks faces decisions about the amount of time and effort that should be invested in recruiting and screening, or possible changes in the conditions of work to consider in order to attract, workers with the desired degree of education, experience, and other qualities. Managers we spoke to often noted that they had a need standard, saying the equivalent of "We don't really have to attract the possibly idiosyncratic superstar; to do our work well, of course we need a solid performer, but here's what we mean." In some other situations, success in extremely challenging work at some frontier may be attainable only with the world's foremost experts. The general point is that the requirements of the job, within the context of the goals and plans of the organization, define a second standard against which a person's characteristics can be judged. This viewpoint seems important for both understanding and action. Knowledge of individuals' capacities is interesting, but is limited in interpretability without this contextual reference point. An effective policy action would seem to be more likely when directed towards a specific target of obtaining the needed quality, rather than towards a vague goal of higher and higher levels of quality.

Although we began our definition by excluding the context in which the employee served, we concluded that this was unwise. First, the question of needed quality provides information about reality that corrects and completes the information about generic capabilities. Second, we think the basic concept of needed quality is important to include so that experience with its measurement can accumulate and strategic judgments about needed human resources can be refined. In light of the uncertainty noted earlier in the chapter, about the precise relationship between individual attributes and eventual government performance, the proposed indicators of both absolute and needed quality should be considered provisional, subject to revision as empirical evidence shows particular indicators to be especially critical.

Specific Indicators of Quality

To begin to narrow our examination of alternative ways to collect data on the two-part definition we propose to use, we divided individual capacities into two areas: knowledge, skills, and abilities, and attitudes, values, and motivation. For these two, and the second part of the definition (the match of individuals' capacities with the job), we listed a number of possible types of data.

Evaluation of Types of Data

Each type of data can be evaluated on two dimensions, the clarity of interpretation about the quality of the workforce that could come from such data and the feasibility of obtaining the data. Measures that seemed likely to result in ambiguous interpretations or that would be very difficult to obtain on the geographically dispersed federal employee group are not proposed for further active consideration at this point; the others are candidates for further exploration. (The analysis here is general; the specific relevance of parts of the data to different analyses of the workforce is discussed in chapters 4-6 below. For example, education indicators are most relevant for those new to the workforce and without work experience.)

We concluded that direct tests of knowledge, skill, and ability (for example, using paper and pencil examinations or observer ratings of individuals in simulated task situations) are not feasible, despite the strong data they provide, because they require controlled testing conditions that are impractical on large dispersed samples. A wide variety of less direct indicators, such as various details of education and work experience, are commonly discussed, however, as signals of quality. For example, a workforce may be considered to have greater knowledge and skill with members who have:

- more formal education;
- better educational results as shown by higher grades, test scores, or class rank;
- more professional training, certification, or licenses; or
- more general or specialized work experience.

We retained this category of data for further exploration.

In the area of attitudes, values, and motivation, there are many possible topics on which employees could report their views (such as, for example, morale, commitment, or job satisfaction). However, interpreting them is less clear than with other kinds of indicators. Because these topics are widely discussed, however, we decided to retain the attitude type of data for further consideration, but chiefly to look for existing data sources (at agency level or government-wide) that would be adequate. If there were few such data, we would have doubts about investing much in new data-gathering of this kind.

Attitudes and motivation towards work may reflect underlying personality dimensions and there are many measurement instruments and methods in that area. There is a historical body of literature discussing "types" of individuals attracted to, or produced by, long-term work in bureaucratic settings in government and elsewhere (for example, persons more interested in job security than in taking risks or in major achievements). Evaluating personality characteristics is, however, a complex measurement problem and interpretation of the results is controversial. We did not consider that type of measurement further for the initial assessment design.

To assess the match of workforce quality to needed quality, it could be desirable to observe work performed by employees with varied characteristics, to empirically establish the capabilities needed (for acceptable output, for high output, etc.). This task could be part of agency personnel management, but indicators requiring direct observation are not feasible as part of the overall assessment we are considering, in light of the sample sizes needed and the dispersed workforce to be assessed.[6] We propose that the initial design begin with data from employee self-reports, supplemented with a second viewpoint obtained from each employee's supervisor.

Thus, after our evaluation, we retained three types of measures for further exploration to implement the definition of quality, including (1) indirect measures of knowledge, skill, and ability; (2) general attitude measures; and (3) assessments by employees and supervisors of the match of personal capacities and the job. (Table 2.1 summarizes our evaluation of various measures and our initial conclusions about each type.)

Table 2.1: Our Evaluation of Types of Measures for Three Dimensions of Workforce Quality

Dimension and measures	Clarity of interpretation	Feasibility of measurement	Initial conclusion
Knowledge, skills, and abilities			
Direct tests	High	Low	Do not attempt
Indirect measures	Medium	High	Explore further
Attitudes, values, and motivation			
Surveys of employee views	Medium	High	Avoid major investment; use existing data if possible
Measurement of basic personality dimensions	Low	Low	Do not attempt
Match of individual capacities and job needs			
Direct observations	High	Low	Do not attempt
Self-assessment	Medium	High	Explore further
Supervisory assessment	Medium	High	Explore further

Specific Indicators Searched for in Agency Employee Data

The three types of potentially useful data shown in table 2.1 can take many specific forms. To check which of these are available in various sources, we developed a comprehensive list of indirect indicators for the knowledge, skill, and ability area, and general questions and probes in the other two areas. We used these as the core items we looked for in our review of agency personnel records, described in the next chapter. (See table 2.2.)

---

[6] The difficulties of establishing workforce needs are sizable. GAO reviews of military manpower planning found the services have difficulties in establishing the basic numbers of workers needed, assuming uniform capability. Unsupported judgments were a chief source of the estimates, although GAO believed hard data were feasible to gather on many of the jobs reviewed. It seems likely to be even more difficult to empirically link different degrees of employee capability to differences in output. See U.S. General Accounting Office, Marine Corps Manpower: Improvements Needed in Processes for Determining Manpower Requirements, GAO/NSIAD-87-102 (Washington, D.C.: May 1987) and Air Force Manpower Program: Improvements Needed in Procedures and Controls, GAO/NSIAD-87-137 (Washington, D.C.: June 1987).

Table 2.2: Potential Indicators for Different Parts of the Definition of Workforce Quality

Dimension	Indicators
Knowledge, skills, and abilities	Education
	Years of schooling
	Degree(s) awarded
	Date of degree(s)
	Grade-point average
	Rank in class
	Institution attended
	Major field of study
	College entrance test scores
	Other test scores
	Continuing education, training
	Quantity (hours, days, units, credits)
	Kind (course names)
	Source
	Professional certificates, licenses
	Examination scores (CPA, bar)
	Certification, licensure record
	Work experience
	General work experience
	Specialized work experience
	Promotion history
	Awards (monetary and other)
	Disciplinary actions
	Performance appraisal score
	Other managerial ratings
	Any other measure or indicator, either generally maintained on record or maintained for specific occupations
Attitudes, values, and motivation	(Open-ended probe for any existing agency survey or other data in this area)
Match of individual capacities and job needs	(Open ended probe for any data used by agencies in this area)

Summary and Conclusions

Because the definitions and concepts used by different observers in judging the federal workforce are so numerous and complex, some priorities must be set as to the most useful definition for a beginning effort at assessment. As experience with an initial data set accumulates, additional measures and further questions can be attempted. This chapter has outlined the following points.

First, of the many possible definitions of quality, a beginning assessment should focus on the individual and his or her work, excluding (leaving for later study) personnel processes, management and the work context, work outcomes, and the broader environment in which federal personnel management operates. Assessment of performance and outcomes is of the greatest importance for improving government, but is a serious measurement challenge and does not respond directly to the committee's request as it would not provide clear indicators concerning the workforce. Data on individual attributes of those in the workforce need to be interpreted cautiously, because of uncertainties in the relationship between these qualities and work results.

Second, once having defined quality in terms of attributes of individuals in the workforce, the definition should include two basic elements: that is, both strong capabilities and the degree of capability needed for the work. This definition emphasizes that individual characteristics should not be judged in isolation and underscores the long-term goal of linking quality assessment to strategic personnel decisions about the needed workforce.

Third, there are types of data that are both generally feasible to obtain and relatively straightforward to interpret that measure aspects of the proposed definition. These include a variety of indicators of knowledge, skill, and ability, as well as measures of attitudes. The match of individuals to their jobs can be assessed through the views of both employees and supervisors. In time, more direct measures in each area can be explored, along with exploring their causal links to work outcomes.

Using Administrative Data to Assess Workforce Quality

Because federal agencies develop a large amount of information on the workforce in the course of routine personnel processes and agency management more generally, we evaluated whether such administrative data could be the source of the information needed for assessing workforce quality. The principal criteria applied in reviewing 8 agencies' records and the central records kept at OPM were comprehensiveness of the quality indicators available and the accessibility and recency of data on them. In addition to providing details of existing data, agencies estimated costs and other factors that would be involved in any attempt to change the existing data system to make workforce data more comprehensive, accessible, or up-to-date.

This chapter is organized in two major parts, the first summarizing case studies of data at 8 agencies, the second reviewing OPM data. Each part presents the data now available as well as the potential for improving what is available.

Data That Are Available in Agency Record Systems

As described in chapter 2, we are proposing that a multi-faceted definition of quality be used, in which three different types of information could be useful. These include indicators of knowledge, skill, and ability; information on attitudes, values, and motivation; and information on the match of individuals' capacities to the needs of a specific job. Eight agencies provided details of records they had on any of these types of measures.[1]

Data on Knowledge, Skill, and Ability Indicators

Table 3.1 shows the availability of 23 items of workforce information that could serve as indicators of one segment of our definition of quality — individuals' knowledge, skill, and ability. About half, 13 of the 23 elements, are stored in the records of 7 or 8 agencies, but adding the criterion of accessibility reduces the total. Only 4 data elements of the list could be easily retrieved from the records of 7 or 8 agencies, and these have significant limitations. One common element is the duration of any training. This item (and the others on training) would be difficult to use because there is less central direction about record-keeping and reporting in this area with the result that agencies define training differently and sometimes store the data in separate data systems apart from the main personnel records. (As discussed in the section on OPM data

---

[1] The eight departments or agencies were: Air Force, Environmental Protection Agency, Health and Human Services, Interior, Justice, National Aeronautics and Space Administration, Navy, and Veterans Administration.

Table 3.1: Number of Agencies Responding Positively to Questions About Workforce Data Elements

	Stored	Easily retrieved	Updated
Education			
Years of schooling	8	5	6[a]
Degree(s) awarded	8	6	6
Date of degree(s)	8	6	6
Major field of study	8	5	6
Institution attended	5	2	1
Grade-point average	1	1	0
Rank in class	0		
College entrance test scores	0		
Other test scores	0		
Continuing education, training			
Source (internal, external)	7	6	6
Duration	8	7[b]	7
Course name	7	6	7
Professional certificates, licenses			
Certification, licensure record	4	2	3
Examination scores (CPA, bar)	2	0	1
Work experience			
General work experience	6	2	3
Specialized work experience	6	2	3
Promotion history	8	7[b]	8
Step increase	8	6	8
Removal actions	8	7[b]	8
Awards (monetary)	8	6	8
Awards (other)	7	5	6
Performance appraisal score	8	8[b]	8
Other managerial ratings	3	2	2

[a] The first four education data elements are updated inconsistently.

[b] Indicates the four most commonly retrievable items, which are discussed in the text.

below, since 1982 OPM only requires agencies to submit aggregated data on training rather than individual-level data.) Two common automated items of possible relevance concern promotions and removals. The fourth common element is a single digit that is the summary score from the agencies' annual performance appraisal. Because rating systems vary and supervisors' standards for making judgments can change from time to time, this data element is not very useful as a general indicator of quality.

Some items listed as "stored" are found in the paper records in individuals' official personnel files, which are decentralized in each agency (in up to hundreds of separate personnel offices per agency, around the country or overseas) and hence difficult to use. Further, few elements meet a criterion of recency; agencies indicated that only the annual performance appraisal score and the payroll-related items such as promotions, step increases, and cash awards were automatically and regularly updated. Agency officials explained that education data elements are not generally updated; updating occurs only at employee initiative or at the time an employee files a new application form when changing jobs. (See table 3.1.)

Data on Attitudes, Values, and Motivation

Employee attitudes form the second of the three parts of the definition of workforce quality. None of the eight agencies reported that recent data from broad employee surveys of this type were included in accessible record systems. Several agencies reported occasionally gathering such data from targeted groups such as managers or workers in a particular unit. Three agencies reported large-scale surveys, but only two of those had been done more than once, and the data were not accessible in any case.

Data on the Match of Individual Capabilities and the Needs of the Job

Agencies match individuals and jobs routinely as part of staffing. However, no information about individuals' degree of capability or the consistency of such capabilities with the needs of the job was reported to be part of accessible record systems by any of the eight agencies surveyed.

Other Information

We also asked agencies if special workforce quality data were kept on 11 specific professional and administrative jobs or whether other indicators of quality were included in the accessible records beyond ones already probed in our survey. With the exception of Air Force and Navy which record some details of employees' job assignments, the agencies responded negatively to both of these questions. We asked whether data were kept on groups of job applicants, to check the possibility of analyzing characteristics of those seeking federal jobs. None of the agencies had systematic information of this kind. In responding to our survey, agencies reviewed their data held at all levels, including headquarters and bureaus, on the chance that quality measurement might be advanced in particular sub-units; several examples came to light, reported in the next section.

Feasibility of Enlarging Administrative Data

The data on existing indicators show that agency practice is diverse concerning the automation of records and that more data are stored than are now automated. In addition, the eight agencies reported idiosyncratic examples of other approaches towards assessment of workforce quality, such as special data items or data sources used by sub-units. For example:

The Department of Justice analyzed applicants for attorney positions by requesting individuals' self-reported law school class rank and any clerkship position (which was evaluated for the prestige or hierarchical level of the court). The data were interpreted as showing the overall quality of applicants, as well as differences in relative quality of applicant groups from different law schools in any particular year. The data are not computerized.

The Environmental Protection Agency maintained special preapplication files on high-quality potential employees in shortage occupations with information gained from recruitment contacts on indicators such as grade-point average and skills.

The U.S. Geological Survey in the Department of the Interior collected annual updated professional resumes from physical scientists who chiefly do research and formally evaluated them using a structured method. The data are not computerized.

The Air Force and Navy code the "skills" of employees as part of the automated record system. (These codes describe details or specializations within the formal job description, for example, that a nurse is assigned to assist in surgery, rather than evaluations or measures of actual skills possessed by an individual.) Air Force (but not Navy) searches the information to automatically generate lists of candidates for vacant positions by matching skill codes in the vacancy notice with skill codes of those eligible to be considered.

The inconsistency and incompleteness of the current automated agency-wide data systems and the individual efforts discussed above both suggest the idea of enlargement of the data bases so that more diverse indicators can be more consistently accessible on a wider population. This might be done by a combination of automating data from the existing paper record or augmenting the source documents and including new data in the automated files. The eight agencies provided information and views on this idea.

Although file space is technically available on at least 7 of the agency data systems, the agencies reported several other potential problems, as shown in table 3.2. Computer programming was most frequently cited as a barrier (by 5 of the 8 agencies) because the older systems in use do not employ up-to-date data base management methods permitting rapid changes. Related problems of editing, testing, and debugging of new programs, needed because of the complexity of changing many instructions, were cited by 3 of the 8. The same number of agencies also reported that there could be a long waiting time for the actual development work, owing to backlogs of planned system enhancements. Other problems cited less frequently but still of potential impact in a wider group of agencies, included implementation problems at highly decentralized data entry points and missing system documentation that impedes some kinds of changes. (See table 3.2.)

Table 3.2: Number of Agencies Citing Barriers to Enlarging Automated Personnel Data Systems

Barrier	Number of agencies
Extensiveness of necessary programming	5
Extensiveness of editing, testing, debugging programs	3
Backlog of system enhancements (long waiting time)	3

Time estimates for preparatory work to add new data elements to the automated systems varied too widely to be interpretable, as did cost estimates, and three agencies did not provide either estimate. Time estimates ranged from one month for a single addition to 6-8 months. Dollar cost estimates ranged from as low as a few hundred dollars to as high as $75,000, despite our efforts to specify the cost elements involved and the scope of the change. Standardizing the estimates by the size of employee population involved yielded per-person costs of the change ranging from $0.01 to $3.33. None of the agencies estimated the additional costs of the time of data entry operators who would need to key information into new data fields in a modified system.

Officials in charge of personnel data at six of the agencies offered several views about the overall concept of assessing quality by monitoring data drawn from administrative records, as outlined in table 3.3. The overall cost was a concern at three agencies; the likely accuracy of the data and the potential for misinterpretation of a relatively limited data set troubled three as well; two mentioned difficulties in gaining agency acceptance of a mandated and burdensome requirement of uncertain local use; and two feared grievances from employees. (See table 3.3.)

Table 3.3: Number of Agencies Citing Specific Concerns About Enlarged Data Systems for Monitoring Quality

Concern	Number of agencies
Potentially high extra costs	3
Inaccurate data; misinterpretation	3
Lack of user acceptance, irrelevance	2
Possibility of employee grievances	2

Data That Are Available in OPM Record Systems

The source of almost all OPM data is the agencies. Therefore, OPM has little additional information beyond that reported in the preceding section of this chapter. We examined the major OPM record system, the Central Personnel Data File, which includes 55 data elements on each of over 2 million federal civilian employees. Each employee's file contains 42 data elements submitted by the agencies and 13 generated by the system; as with agency data, the CPDF elements primarily reflect employees' current status on items related to pay and benefits. The education items, drawn from agency submissions, share the problems of limited overall comprehensiveness and inconsistent updating of the few elements that are on the file. OPM officials believe little can be inferred from these data elements.

We found several other OPM data-collection efforts, but they are inherently limited or have been modified or discontinued so that they offer little at present. These include the following:

Training data on individuals used to be part of the CPDF, but officials at OPM said they were dropped when budgets were reduced in the early 1980s because there were few users of the data. Information on training is now collected from agencies and entered in a separate record system only in aggregated form. It is difficult to interpret because a key item not collected is the course title.

A government-wide attitude survey begun to aid in evaluating the Civil Service Reform Act of 1978 has not been re-administered since the third round in the early 1980s.

The Professional and Administrative Career Examination was discontinued as part of the settlement of a court challenge of its validity for selection. OPM officials could not recall any analyses of score trends in the overall tested population during the period it was in use, which might have been useful if all or parts of the test could be administered to new samples for assessment purposes.

No archive is kept of results of OPM evaluation of various data on applicants (education, work experience, etc.) in those occupations for which central registers are still maintained, so no trend analysis of applicants is possible (except in the small number of job series where alternative tests have been developed).

Modifications in OPM Data

There is very little statutory direction for the nature of personnel data that agencies must report to the central file and OPM officials believe there is little effective sanction for agency non-compliance with data requests other than the hypothetical withdrawal of appointing authority. Thus the CPDF is only as complex as permitted by voluntary cooperation, user needs, and available funds. The most recent overall review of the substantive content of the file was about ten years ago. OPM officials said that over 400 suggestions for augmentation were reviewed, but only 15 were adopted. (A significant redesign completed in 1987 improved technical aspects of the CPDF.)

According to OPM officials, any proposed new data items for the CPDF are reviewed on criteria of legal authority, cost, collectability, privacy considerations, and potential utility. Even when proposals are approved following OPM and agency review, long lead times are to be expected as agencies alter their systems at all levels as needed for the change; basic compliance is piecemeal and data quality is expected to be poor for the first few years.[2]

---

[2] Compliance is incomplete after a year even in a case where OPM requested a new data element for the CPDF from agencies (employee health benefit plan) under presumably favorable conditions. In that case, agencies already had the code automated in their own records, specific OPM offices needed the data for clear policy purposes, and significant federal funds were at issue.

OPM officials encourage alternatives such as sample surveys rather than the complex process of changing CPDF. Various offices within OPM are major users of CPDF and are apparently satisfied. Individual agencies have more extensive data on their own employees and apparently rarely need comparative data. Although there may have been scattered requests, OPM officials familiar with CPDF did not recall any steady demand from the major current users for enhanced information on workforce quality.

Summary and Conclusions

Review of data now kept at eight agencies showed that information on indicators of workforce quality is limited in comprehensiveness, accessibility, and recency. A few measures of education are maintained in central computer files, but these are not consistently updated. Other indicators of education or work experience are available only on paper records in employees' official personnel folders, which are kept at decentralized personnel offices. Nothing on attitudes or job-skill match appeared consistently or in accessible form. The bulk of data that are accessible and up-to-date pertain to pay and benefits, which do not match the items needed for the proposed definition of workforce quality.

Personnel data managers at the eight agencies raised concerns about cost, feasibility, and user acceptance in discussing the possibility of enlarging the existing data systems to include data either from existing paper records or from new sources. Officials were unable to provide consistent and comparable cost estimates for system modifications, but cited backlogs of needed system changes and technical problems associated with any type of change to the common older computer hardware and software in use in the personnel record systems. These officials doubted that the potential use of data on workforce quality would be vigorous enough to assure acceptance of mandates to enlarge the systems.

Data in OPM's major system, the CPDF, include a small number of relevant items (on education) but they are derived from those in the agency systems and are not up-to-date. There are few other kinds of usable information at OPM. Chances of change there appear slim as well.

Since existing sources are inadequate and modifications unlikely, assessment of workforce quality must rely on alternative data sources. Both OPM and agency officials believe sample surveys could be an effective alternative.

Assessing the Entering Workforce

The next three chapters outline our proposal for an ongoing program of assessment of workforce quality. The overall question of quality is to be addressed by a repeated assessment of the current workforce, at intervals, to assess changes, which will be supplemented by periodic studies of those entering and leaving the workforce. For each of these three approaches to studying quality we propose new data-gathering by sample surveys; the three chapters discuss the details of our proposal in each case. This chapter presents the initial segment on assessing the entering workforce.

New staff serve to replenish the lifeblood of an organization and in fiscal year 1987 the federal government hired 40,573 new people for full-time permanent civilian professional and administrative jobs. Assessment of the quality of workers entering the federal workforce is of great interest for judging whether the capabilities available for federal work are being adequately kept up and for use with other data in reviewing policies such as pay and benefits or processes such as recruitment and selection. Questions in this area are increasingly important as professional and administrative jobs expand in significance in the federal sector, which places government more and more in competition with others for employees with advanced knowledge and skill.

This chapter outlines major questions to be answered about the entering workforce, evaluates alternative sources of data, and proposes a basic design.

Questions About the Entering Workforce

Any assessment in this area must answer two basic questions: How good are those attracted to federal jobs and how do they compare with others hired elsewhere? Within the framework of this general question, however, it is important to draw the proper specific comparison. The stereotyped view held by managers and co-workers is that newcomers at any particular time are rarely "as good" as those who are in or are leaving federal jobs. But such a comparison of entering versus current or exiting employees is inappropriate since by definition the newcomers come from a different labor market, may have been selected using different criteria, and will always lack the old-timers' federal job experience. If there were data on newcomers at a series of points going back in time, today's entering workforce could be contrasted properly, not with today's experienced workers but with yesterday's newcomers. By starting the proposed assessment of incoming employees in comparable federal and nonfederal jobs and repeating it, such a series will be established.

The question is not only the general quality of new federal employees, but the quality of the workforce in specific jobs. Although some policies, such as pay and benefits, affect federal jobs in general, most discussions of shortages of needed employee skills, most practical remedial actions, and even the special pay rate process that determines the need for higher salaries in areas of shortage, focus on specific occupations. Therefore data to answer questions about quality need to be at the level of specific occupational series.

Some who reviewed our proposed design believed that the question should be refined even further, to ask whether there are differences in the quality of those entering an occupation in different agencies. The rationale is that for specific occupations, such as attorney or accountant, agencies differ in the type and conditions of work, the labor market conditions they face in recruitment and retention, and the effort made to recruit a strong workforce. Consequently, it is argued, any assessment should provide data that could reflect the results of such differences in the actual quality of each agency's new hires. The analysis is plausible and the question is appropriate, but answering it may be technically difficult. Similar questions could be raised about agency differences in overall workforce quality or in separated employees. Several general problems in reaching the goal of agency-level analyses of workforce quality are discussed in appendix III.

Data Needed to Answer the Questions and Potential Sources

The general problem for designing any data-gathering effort is how to provide sufficiently interpretable information in a feasible, timely, and affordable fashion. A range of sources can be considered, each with its own strengths and weaknesses.

The definition of quality for the proposed assessment specifies two broad areas of information, individual capabilities and the relation of the individual's degree of capability to the needs of a specific job. Though new employees should be matched to their work just as should any other employees, given the way we are proposing to measure that match, strong data may not be available early in an employee's career. That is, because new employees are relatively unfamiliar with the work, data on their perceptions of the match of their skills to the needs of that work could be unreliable. The supervisor would give a more informed view, but depending on the employee's time on the job, even the supervisor may not have had the time to observe a fair test of the match. We are proposing, therefore, that the assessment can be restricted to the first part of the definition, so that information about new employees is needed only on indicators of individual capability. (The second part of the definition is emphasized more in the next chapter, in the assessment of the overall workforce.) For individuals new to the workforce as well as the job, capability can be assessed chiefly through educational indicators. Some will have experience as well, especially those hired at higher levels.

Because the information is needed not only on federal employees newly hired, but on comparison groups of others hired for similar jobs elsewhere at roughly the same time, the question of data sources is complex. There is no source of comprehensive information on the characteristics of people in specific occupations and gathering new data presents significant sampling challenges.

The 3 X 3 matrix in table 4.1 summarizes our assessment of the performance of several sources of data on employees' education and experience with regard to feasibility and interpretability. We define feasibility as the degree of technical effort and cost needed for such tasks as drawing samples, obtaining access to existing data, or collecting new data. Interpretability refers to the clarity of inferences that can be drawn from the data on the questions of workforce quality. (See table 4.1.)

Table 4.1: The Feasibility and Interpretability of Alternative Data Sources for Assessing New Employees

	Interpretability		
Feasibility	**Low**	**Medium**	**High**
Low			School records on graduates' achievement and employment Employers' files on employees
Medium	Employers' opinions on education and work background of employees	Employees' self-reports (nonfederal)	
High	School faculty or administrators' opinions on graduates' achievement and employment	Employees' self-reports (federal) College entrance test scores	

The diagonal pattern in the table suggests the general conclusion that the easier information to obtain is less useful and the most interpretable data are the most difficult to obtain. For example, the upper right segment of the table shows a potentially useful source that we found is not available. If schools maintained records that would allow linking students' school achievement and their later job (both occupation and employer), a sample of graduates could be drawn, the records could be searched to extract these data, and analysis would involve comparing those in federal jobs with others. However, when we discussed their records with officials at a number of colleges and universities, few had systematic information on all types of graduates (some schools track some kinds of graduates) and academic achievement records are maintained separately from graduates' placement records.

New employees could be located at their work, where employer records such as job applications, interviews, or tests, might provide the needed data on employee education and experience. However, the extensive information submitted by federal job applicants (on a common application form known as standard form 171) is not entered into accessible computer files but kept in the official personnel folder, so that extraction of the data for a sample would require searching in hundreds of scattered personnel offices. (Federal hiring is increasingly decentralized; for the small number of occupations for which central hiring registers are maintained, data resulting from the centralized evaluation of applicants is discarded routinely.) Access to employer files outside the federal government is uncertain and the information would vary in extent, comprehensiveness, and format.

In contrast to the difficulties of locating these data from formal records on individuals, it would be much easier to gather opinion survey data in which knowledgeable people characterize groups. For example, either at schools or work places, officials (placement officers, faculty, personnel specialists, general managers) could provide views on characteristics such as education, work experience, and attitudes of people entering various occupations in federal and nonfederal settings. Indeed, this has been done several times recently with federal managers and personnel officials, as cited in chapter 1, and the Merit Systems Protection Board surveyed higher education officials' views several years ago also. Even students themselves, before choosing jobs, could provide opinions about their characteristics and how they seem to be matched to different places of work. However, no matter how systematically the opinion survey might be done, the resulting data would be hard to interpret. Nothing new would be provided by repeating such surveys and our objective is to go beyond the shortcomings of such opinion data.

Intermediate in interpretability are two kinds of data that we propose for use here, self-reports and college entrance test scores; both are quite feasible. First, individuals new to their jobs (both federal and other jobs) could provide data on their education and work background. Direct questioning of individuals avoids the problems of access and comparability that arise in using existing records. The data obtained from self-reports can be inaccurate (from weak memory or deliberate distortion), but the amount of bias may be small since the self-report is not for decisions but for research. In any case, the self-reported data for a sub-sample would be checked against other sources at an early point in the assessment project to estimate the degree of bias. School attended tells little by itself, but can be interpreted as a secondary indicator about the individual by use of a school quality measure.[1] Feasibility is high for locating new federal employees to gather the needed data, but locating samples of nonfederal employees in specific occupations is a challenge.

National standardized college entrance test scores are also of only moderate interpretability, shown by the long history of controversy about their meaning, though they continue to be used as aggregate indicators of educational outcomes.[2] However, we determined that it is feasible to obtain the scores by contracting with college entrance testing organizations for computer matching. Upon official approval by the authorities who control the examination data on behalf of the original students, the firms can extract from their historical files any scores for members of study samples based on name, date of birth, and social security number. To preserve confidentiality, the data are returned for analysis in a way that keeps individuals' names and scores from being associated. For federal professional jobs for which there are specific educational requirements and most new employees have college degrees, such scores would probably be on file for most of those in any sample; the same is true for any nonfederal group as well. For the group entering administrative jobs for which college is not required there would be some missing data. The scores from the two different tests cannot be combined. The feasibility of obtaining and interpreting other standardized test data in specific fields (for example, tests in fields such as law, medicine, nursing, or accounting) can be explored when specific job series are chosen for assessment.

Proposed Design

The basic design is to survey samples of recently hired federal workers in selected occupations and comparison groups of nonfederal workers recently hired in similar jobs. The next sections discuss the occupations which could be selected, the sampling approaches, and the indicators on which data would be gathered.

---

[1] There are several existing rankings of higher education institutions or units within institutions, using measures from various sources (empirical data or informed judgments) on faculty and students. See David S. Webster, Academic Quality Rankings of American Colleges and Universities (Springfield, Ill.: Charles C. Thomas, 1986).

[2] The Department of Education gives nationwide prominence to statewide average scores on the two common college entrance tests (the Scholastic Aptitude Test of the College Entrance Examination Board and the exam set by the American College Testing Program) in an annual "wall chart" of educational indicators.

Occupations to Be Selected

This assessment should focus on occupations for which entry quality is of particular interest. Various criteria could be used to select jobs for study, such as unusual losses of experienced employees, particularly acute pay disparities that might have caused lowered recruitment success, or anticipated changes of mission that are believed likely to increase the importance of the quality of employees being hired. Assessing a number of occupations at once allows for efficiencies at all stages of the work, though there is nothing particular about measurement or analysis that dictates how many occupations are assessed or when. For purposes of estimating the overall costs of an initial assessment, we have assumed that 12 occupations would be selected and that samples of federal and nonfederal employees in all of them would be surveyed at the same time (a total sample size of about 4,800). This arrangement would permit analysis of a sizable body of information, which seems preferable since there is likely to be great interest in the results. The alternative of spreading occupations across several years would delay providing the full results. Appendix III gives further details of the sample size and costs under these assumptions.[3]

Sampling of New Federal Employees

A sample of newly-hired federal employees in each occupational series of interest can be readily drawn using the Central Personnel Data File and surveys can be distributed to these individuals through their personnel office, which is also shown on the CPDF record. We discussed with agency officials the feasibility of reaching people sampled in this way, since agencies have experienced the method as used in surveys from OPM and MSPB in recent years. Although some degree of burden is inevitable, the method poses no special difficulties.

---

[3] The four-year interval between measurement cycles suggested for the core assessment may be reasonable here as well, except that places a large work load on the staff to mount both at the same time. Some key occupations should be assessed again, especially if policy changes have affected them. But others can be added.

Sampling of the Nonfederal Comparison Group

We located two approaches to sampling the nonfederal group, through employers or through higher education institutions.[4] The first way is to use lists of employers developed as part of state unemployment insurance record systems. From these, and information on where specific occupations are concentrated in different types of firms, it would be possible to sample employers likely to have hired people in jobs comparable to the chosen federal occupational series.[5] The numbers of new hires will vary greatly by employer so sampling may be necessary within some firms. In general, we conceived that the next step would take one of three forms. Employers could provide lists of all newly-hired people in the occupations for the assessment staff to use in sending out surveys (to all or a sample). Alternatively, the employer could distribute surveys directly to selected new employees (to be returned directly to the federal agency administering the assessment). In a third alternative, the employer could extract data from individual employees' files.

We attempted to more precisely evaluate the feasibility and costs of these alternatives with experts in the Bureau of Labor Statistics (BLS) who work with states in surveying employers and who also use a national merged employer list routinely for gathering aggregated data on numbers of people working in specific jobs. We learned that BLS has no experience using the employer lists to try to contact or gather data on individuals as our design requires. Thus BLS officials could not evaluate the feasibility or cost of obtaining the new-employee information. A pilot test of one or more of the three approaches would, however, answer such questions readily. For efficiency, this methodology could be attempted in only a few states. These could be chosen on criteria such as high existing concentration of federal employees in the target occupations or exceptionally keen competition for entry-level employees in the selected occupations. Though its feasibility is not clear, the employer approach appears to deserve further study because it would yield results on a generalizable population of new hires at all levels of experience working for a wide variety of employers including state and local government, nonprofit institutions, as well as private firms.

---

[4] See appendix III for more details of this sampling problem.

[5] An important criterion is similarity of work, since proper interpretation of data on new employees' capability requires that work in federal and non-federal jobs with the same title be equivalent. The comparability of work under various systems of job classification has been examined in detail by the Bureau of Labor Statistics and OPM to produce "cross-walks" from the more general Dictionary of Occupational Titles to the unique job titles in the classification system used in the Federal Personnel Manual.

Because of the uncertainty of locating new nonfederal employees via their employers, we examined a second approach involving surveying recent college graduates sampled and located from school records. We found the Department of Education periodically conducted a Recent College Graduates Survey that sampled those receiving either bachelor's or master's degrees and questioned them approximately one year after the degree date. We examined whether this survey, with modification, could be used for the proposed quality assessment, but as we finished our work the department decided not to continue this survey. The department is planning a more complex longitudinal survey of college students to begin in 1990, which could also serve our purposes. Appendix III gives further details of our review of the prior survey and the issues that would need to be resolved if the planned future survey is similar. (The main concerns include augmenting the survey questions so that data are gathered on our quality indicators including details of education and work experience and assuring that enough people are located in the specific occupations selected for assessment.)

There are two drawbacks to a design based on using any survey of recent graduates. First, it provides a comparison group only of those entering jobs shortly after obtaining a degree. The federal employee sample could be made comparable, but the assessment in that case would not provide data on qualities of more experienced workers entering the federal and nonfederal sectors. Second, for those jobs where a degree is not required (such as all the federal jobs categorized as administrative), having a comparison group of degree-holders only will overstate the education levels in comparable nonfederal jobs. Nevertheless, this approach can be explored further, as it appears to require only modest modifications to a planned survey, if the preferable alternative approach through employers turns out to be impractical.

Indicators

Table 4.2 shows the data elements from the general set that are needed for this particular assessment. All but the test scores would be gathered by surveys: of individual employees in the case of the federal workforce and of either individual employees or their employers in the case of the nonfederal workforce comparison group (as discussed above and in appendix III). (See table 4.2.)

Table 4.2: Dimensions, Indicators, and Data Sources for Assessing New Employees

Dimension and indicators	Source of data
Knowledge, skill, ability	
Education	
Years of schooling	Employee survey
Degree(s) awarded	" "
Date of degree(s)	" "
Grade-point average	" "
Rank in class	" "
Institution attended	" "
Major field of study	" "
College entrance test scores	Testing organizations
Other test scores	(to be determined)
Continuing education, training[a]	
Quantity (hours, days, units, credits)	Employee survey
Kind (course names)	" "
Source	" "
Professional certificates, licenses	
Examination scores (CPA, bar)	Employee survey
Certification, licensure record	" "
Work experience[a]	
General work experience	Employee survey
Specialized work experience	" "
Promotion history	" "
Awards (monetary and other)	" "
Attitudes, values, and motivation	Employee survey
Match of individual capacities and job needs	Not Applicable

[a] Applicable to recently-hired employees with prior work experience.

The table lists 9 education indicators; all are commonly used to characterize the kind, quantity, or quality of schooling, but the list is almost certainly redundant to some degree. As experience with the assessment accumulates the list can be refined and the best data items retained. The category of other tests could include, among others, graduate entrance examinations of various kinds, such as the Law School Admissions Test; their accessibility and interpretability can be explored when specific occupations are selected for study. As the movement to measure outcomes in higher education grows, there may be more and more test scores on sizable populations, such as all graduates in large state systems, that could be explored for this assessment. Work experience and continuing education are relevant indicators for those entering a new job with some background, but the feasibility of comparison data on nonfederal new-hires of this type is somewhat uncertain, as discussed above. Performance appraisals of employees from a wide variety of previous employers may be difficult to interpret, so we do not suggest these be collected. Extensive new attitude survey data generally are not proposed for this assessment, but a few questions could be included in the employee surveys, for example, to gather reasons for taking different jobs, evaluations of recruitment processes, or commitment to the employer.

The analysis of the data would be straightforward, examining differences between new federal and nonfederal employees (within the same occupation) in the degrees earned, grade point averages, class standing, test scores, institutional quality ratings, years of experience, and so forth. Major pitfalls in analyzing and using these data could be avoided by taking specific steps such as:

- warning users of the importance of judging individuals on their own performance and not attributing to any individual the aggregate characteristics of a group;
- checking results carefully for significance; avoiding overemphasis on small differences in relatively crude measurements such as self-reported grade-point-averages;
- releasing data on a wide range of indicators at once to allow the most balanced and comprehensive interpretation and to avoid undue publicity for single indicators such as college entrance test score; and
- avoiding any generalization beyond the specific occupational series selected for study.

Summary and Conclusions

Probably the largest amount of discussion of the quality of the federal workforce centers on those being hired. For this group interpretable measures of quality are feasible to obtain, including education, work experience, and test scores. The critical analysis will be to compare characteristics of new federal and nonfederal employees in similar jobs. Some effort will be required to obtain the needed nonfederal samples and data. Indicators such as education and test scores do not provide direct judgments of how well people will do on the job, but they do provide useful relative indicators about those entering federal and nonfederal jobs and can signal comparative problems that deserve closer study and action.

With the recommended study design, data would be available for the first time to help provide generalized answers to such questions as:

- To what degree does the federal government attract new professional employees, such as lawyers and accountants, from highly-regarded colleges and universities?
- Do those entering federal contracting positions have less training and experience than their counterparts in private industry?
- What is the academic standing of incoming federal professionals compared to that of people hired in other sectors and are there some occupations that are better or worse than others in this respect?
- In an occupation like nursing in which many different degrees and credentials are possible, do those entering federal service have patterns of degrees and credentials different from those taking jobs elsewhere?

Chapter 5

Assessing Changes in the Workforce

The federal workforce is a dynamic system in which change is continuous. But what are the cumulative effects of the myriad changes? On balance, are employees better-educated now than earlier? Is their training equipping them to keep pace with their changing work? To highlight important trends such as these, it is useful to freeze the system and draw a portrait, which can then be compared to portraits from the same vantage point at intervals later. Changes should stand out clearly when the images are juxtaposed. Designing the database for this long-term analysis has been the main focus of our work.

When considering the entire federal professional and administrative workforce, made up of people with a wide range of experience, questions about quality should not focus on initial qualifications alone. Trends in education indicators such as those emphasized in the previous chapter may indicate short-term advantages or disadvantages of federal employers in the entry-level labor market. For this part of the assessment, however, the emphasis should be on broader questions of the extent of capacities developed in the workforce through all kinds of education, training, and experience. Further, we should try to find out whether such capacities are being properly applied to the work, so that the workforce is neither over- nor under-qualified for its assignments.

This chapter presents the design for the second proposed segment of the assessment, to provide a broad picture of the workforce, intended to be repeated at planned intervals. Sections of the chapter present major questions guiding the design, data needed, evaluation of alternative data sources, and specifics of the design including occupations to be studied, sampling, and indicators.

Questions About the General Workforce

The basic question here is how good is the workforce and how has that changed since the last look at it? The question can in fact be posed quite generally and a general answer could be meaningful with statements about the level of education that prevails in the workforce at a certain point, for instance. Because of the growing predominance of white-collar jobs in government (excluding the Postal Service), the federal workforce in recent years has had a higher education level than the overall workforce. A sizable number of occupations should be included so that such general statements are well-grounded.

But it is more probable that questions will center on comparisons within specific occupations as well, so the design should not aim simply to be roughly representative of the professional and administrative group, but must also be targeted on specific occupational series of interest. Scientists and engineers, for instance, will undoubtedly continue to be of interest. As in the first question, addressed in chapter 4, on the quality of the entering workforce, questions will also be asked about agency differences, which may be feasible to answer about some occupations but not others. So that the design can assure an initial consistent baseline of information across all the jobs examined, we suggest postponing attempts to provide agency-level comparisons, though after further technical development, that is a logical candidate for an expansion of the effort.[1]

The question of quality in the current general workforce (as opposed to those just entering) does involve the second part of our overall definition of quality. That is, it is important and logical here to inquire if the workforce is well-matched to its work. In addition, attitudes, values, and motivation toward the work can be examined, given that this population (again in contrast to newcomers) has a more reasonable basis for making statements on such subjects.

Answers to questions about quality should take into consideration the employee's length of service, so that patterns can be identified in broad subgroups of the workforce who entered at the same periods, in addition to those in the same jobs. This distinction, pursued over time, allows answering questions, for instance, about whether the experienced workforce is as good as it was at earlier points in time.

Data Needed to Answer the Questions and Potential Sources

We considered a number of sources which might provide the data needed on the various segments of the definition of quality. The 3 X 3 matrix in table 5.1 summarizes our assessment of the performance of several sources of data on the same criteria discussed in the previous chapter: feasibility of gathering the data and interpretability. (See table 5.1.)

---

[1] The main issues are whether, as the desired analyses expand, there are enough employees in the various categories in which comparison is wanted for usable samples to be drawn and the added cost of enlarging samples where feasible. Appendix III includes discussion of the problem of agency-level sampling and analysis.

Table 5.1: The Feasibility and Interpretability of Alternative Data Sources for Assessing the General Workforce

	Interpretability		
Feasibility	**Low**	**Medium**	**High**
Low	Employee data in official records	Behavioral simulations evaluated by experts	Direct tests of employee knowledge, skill, and ability Direct observation of work
Medium			
High	Supervisors' opinion of employees' characteristics, attitudes Behavioral data, unobtrusive indicators (e.g. leave usage) College entrance test scores	Employees' self-reports of characteristics, attitudes, and job match Supervisors' report on employee skill/job match	

Direct measures (tests, observations, or simulation exercises) as discussed in chapter 2 in general, are not feasible for the start of this assessment though they could be experimented with as development proceeds. If common data were gathered on employees' performance in standardized simulation exercises used, for instance, in various agency executive development programs covering common issues such as supervision or group leadership, over time a sizable body of data would accumulate. Data on employees in their official files, as discussed in chapter 3, have the disadvantages of being both limited and inaccessible.

There are a number of feasible alternatives, which vary in usefulness. Lowest in interpretability would be data from surveys of supervisors' opinions about their employees' background, such as education. Supervisors are not useful sources of such factual data nor probably of information on employee attitudes. Employees leave traces in formal records, such as their use of leave. But the interpretation of such data in isolation is impossible. Scores on college entrance examinations taken years earlier, though easy to retrieve, offer little useful information on the general workforce. (We considered such scores somewhat more interpretable with respect to the entering workforce because of the lack of other indicators on that group and the likelihood that the tests had been taken more recently.)

Useful and interpretable data could come from asking employees to report facts about their own background (education, work history, and so forth), to report their attitudes, and to evaluate the match of their capacities to their work. Surveyed for research, outside of a formal decision context, employees would probably give fairly accurate information.

Because of the importance of the job-match aspect of the definition in assessing the overall workforce, it is crucial to have a second viewpoint, that of the supervisor. Though this adds to the costs of the assessment by doubling the sample size in this segment, the added value of the additional perspective, we feel, justifies it.

Proposed Design

Occupations to Be Selected

The important criterion for judging the sample is coverage of diverse occupational series across the spectrum of professional and administrative work. The exact number is not critical. Generalization to very large populations (such as all professional employees) is not an objective; accordingly, statistical precision is only an issue within each occupational series, where enough members should be sampled to provide conclusions at an acceptable level of likely error.

Review of the overall list of federal professional and administrative jobs showed that 23 occupational series included slightly more than half the total group. Appendix II gives more detail of the specific series we recommend be included in the assessment. Two of the 23 are clusters of physical and biological science jobs.

As repeated data collection is the requirement for trend analyses some core of occupations should remain stable from one assessment to the next. Additions and subtractions in the list, however, pose no problems. The assessment should be repeated every four years, to provide a data point between each presidential election.

Sampling

The sample of employees would be drawn from the master list in the CPDF which includes most (though not all) federal employees. Within each occupation, we suggest that the workforce be grouped into three clusters (0 to 3 years of experience, 4 to 10, and more than 10) and random samples drawn from each. These samples would be drawn every time the survey is administered, such as at the four-year intervals we recommend. We considered the additional information that could be gained by using a panel design, in which an initial sample drawn for one assessment would be recontacted at the time of a second assessment. The strength of panel designs is that by tracing the same people across time they allow much better analysis of what may have caused changes in the group. We decided not to suggest a panel design, however, because we believe the assessment should begin by providing a set of descriptive data only. Even if we aimed for causal analysis, we would want to explore whether there may be more cost-effective alternatives, because of the complex logistics of maintaining the panel in a sampling design that has several strata of occupations and levels of experience as ours does. Extensive work would probably be needed to trace individuals and accurately replace those who leave the different segments of the sample.

The overall sample would include about 27,600 employees and an equal number of supervisors matched to each employee. (See appendix III for details concerning sample sizes.) The employee sample size is comparable in scale to OPM and MSPB surveys of the workforce done in recent years.

The CPDF record gives the employee's location in government, including the personnel office that maintains the official record. The supervisors for each employee sampled would be determined by the personnel offices. Surveys would be sent to the appropriate personnel offices to be forwarded to the supervisors of record to be filled out on each of the sampled individuals.

We do not propose a comparison group of current employees in the private sector for this part of the assessment as we did in the previous chapter on entering employees. Our analysis of questions about the overall federal workforce showed that the more frequent concern is over changes within the federal group over time. In addition, as emphasized in chapter 4, there are significant problems of obtaining samples of nonfederal employees for new data-gathering, and we found no existing sources of data on our indicators of quality for a broad group of nonfederal occupations.

Indicators

Table 5.2 shows the data elements for this segment of the assessment and their sources. The employee and supervisor are the sources for all the new data proposed. (See table 5.2.)

Table 5.2: Dimensions, Indicators, and Data Sources for Assessing the General Workforce

Dimension and indicators	Source of data
Knowledge, skill, ability	
Education	
Years of schooling	Employee survey
Degree(s) awarded	" "
Date of degree(s)	" "
Grade-point average	" "
Rank in class	" "
Institution attended	" "
Major field of study	" "
Continuing education, training	
Quantity (hours, days, units, credits)	" "
Kind (course names)	" "
Source	" "
Professional certificates, licenses	
Examination scores (CPA, bar)	" "
Certification, licensure record	" "
Work experience	
General work experience	" "
Specialized work experience	" "
Promotion history	" "
Awards (monetary and other)	" "
Attitudes, values, and motivation	Existing employee survey done by MSPB (with possible modification as needed)
Match of individual capacities and job needs	Employee self-appraisal Supervisory appraisal

As noted in chapter 4, the education indicators can be refined as experience with their use accumulates. Work experience is an important indicator and experimentation will be needed to determine the best format for efficient surveying of needed information. Experience with the assessment will also suggest data elements of greatest power and interest within the work experience area, among such aspects as the positions held, moves made, rewards obtained, or others. There is very wide experience in personnel with various ways of structuring questions on biographical items like these. Training is an important indicator for those who are some years beyond formal education. There are no central data gathered on individual federal employees' training at present. Thus some development will be needed to specify the questions in this area also so that employees can accurately report on the kind and extent of training and their achievement.

Attitudes and related views could certainly be gathered as part of the employee survey instrument although there is a sizable continuing assessment of federal employee attitudes which, with modification, may provide sufficient information. The Merit Systems Protection Board periodically conducts its Merit Principles Survey, most recently sampling 21,000 federal employees and asking a number of general questions about morale and job satisfaction, among other subjects. There is no statutory requirement that such a survey be done regularly, so if it (or an augmented version) does seem usable in the context of the assessment of quality, it would be prudent to provide more assurance that the data series would continue. Because the previous MSPB surveys have been designed for somewhat different purposes, one design issue would need to be reviewed and possibly some adjustments made. Past MSPB employee samples have covered the entire workforce (not just professional and administrative workers) and have allowed analysis by agency and also by grade levels government-wide, but not by occupation. Since occupation is one critical dimension in our design, some augmentation of the MSPB sample might be needed to insure adequate samples sizes for purposes of the proposed quality assessment.

Because of plausible links of employee turnover and productivity to attitudes such as morale, job satisfaction, or commitment to the present job, measures of these employee views are sometimes discussed as part of analysis of the quality of the workforce. However, data in news articles or popular discussion may be based on poorly drafted questions and unrepresentative samples. If such employee views are to be included in an assessment, we believe they should be measured as regularly and soundly as possible and there is substantial experience with technical issues in the field to draw on both in social science attitude measurement and in public opinion polling. As indicated in our discussion of the definition of quality in chapter 2, this type of data requires cautious interpretation because of uncertainty about how the attitudes being measured are related to behavior. For example, does employee morale cause work effort or does it reflect external events unrelated to work? All measures of workforce attributes have a degree of uncertainty as to their link to other variables but we emphasize a special word of caution in this regard about the often-cited attitude variables.

The match of qualifications and job can be examined through rating scales completed by both employee and supervisor. The instrument would differ from the familiar managerial appraisal forms by providing much more specific dimensions for rating, though it still need not be lengthy. Items could include some or all of those included in the 2-page research performance appraisal instrument developed by OPM and used as the dependent or criterion measure against which to test various predictors of work success. That instrument covered general dimensions such as the employee's work output, quality of work, accuracy of work, job knowledge, and diversity of tasks that can be done by the employee. It also requested supervisors' separate ratings of both the importance of specific abilities and the employee's level of each, such as oral expression, written expression, reasoning, work relations with others, or adapting to changes in work. We believe the employee form should require substantiation of the self-assessment ratings (citations of work experiences or training, for example) to put some pressure on the employee for accurate responses. OPM has experience with such instruments and has found the administration feasible and the results interpretable. Further technical development can determine the proper mix of general and specific abilities to be rated.

Summary and Conclusions

Useful data can be obtained to assess changes in the quality of the professional and administrative workforce. The two key dimensions include knowledge, skill, and ability; and the match of the individual's degree of capability to the needs of the job. Major data sources are the individual and the supervisor. A current government-wide employee survey done by MSPB may, with modification, serve the needs of the assessment for attitude and morale indicators.

With proper design of repeated measures, data would be available for the first time to help provide generalized answers to questions such as:

- Are today's experienced employees in the selected attorney and accountant series as good as those of 4 or 8 (or more) years ago?
- Are we getting the skills we need in the workforce of contract specialists? Has that changed?
- How good are the scientists (in the selected series) who stay 10 years in government, compared to earlier times?

Chapter 6

Assessing Separations From the Workforce

The other side of the coin from attracting a quality workforce is whether the best are retained. Retention needs to be long enough so that recruitment, orientation, and later training costs are recouped. It should also be targeted enough so that the most desirable employees stay on the job the longest. Those who leave may take with them leadership and management skills developed over long periods and especially suited to the special situations of federal programs. In areas in which government is working to develop greater expertise than it now has, such as use of computers, progress may be especially slowed if capable people are lost to other sectors. Where government may be in an adversarial role involving oversight or litigation with other sectors, in areas such as antitrust, the environment, or contracting and procurement, loss of expertise to the other sectors can affect the balance in the relationship. Speculations and anecdotes about the quality of those leaving government are common, so that an ability to measure this aspect of changing quality also became an important focus for our work.

This chapter completes the design of an overall framework for assessing the quality of the workforce by giving details of an approach to evaluating the quality of those leaving federal jobs. The chapter has three sections, following the outline of the two preceding design chapters: the first examines questions about the subject; the second reviews data needed to answer the questions and possible sources; the third gives details of the proposed design including approaches to selecting occupations that might be studied, sampling considerations, and indicators.

Questions About Those Leaving the Workforce

The basic question here is whether the federal government is losing its best employees. If the answer is yes, that suggests a need to strengthen the incentives that maintain employees' commitment to their federal work. The overall assessment outlined in the preceding chapter is intended to answer general questions about the workforce across time and whether it is suited to its tasks. In a general sense, merely by examining the trends in quality within the three levels of experience in the workforce sampled in such a study, the question of how well the government is holding on to quality employees could be tentatively answered. But trends in quality indicators in an aggregate group can result from both incoming employees as well as exiting ones, so that answer would not be as accurate or conclusive as might be wished about those being lost. More importantly, answers take a long time to develop in trend studies. The interval of observation we are suggesting is 4 years, so even if a major baseline study were done immediately, the trend data would not be available for about 5 years. Therefore, more immediate information would be useful, especially by providing details about the characteristics of those leaving that will not in any case be available from the core assessment, which examines quality only in the workforce remaining.

To know whether separations include disproportionate numbers of the most highly desirable employees, it is necessary to compare those leaving and those staying, to see if the two groups differ in quality. Questions focused on the first part of our definition of quality presented in chapter 2, concerning capabilities of individuals, are most relevant to this assessment. That is, it is of most interest to know if, on average, the separated employees are better than the rest in their degree of knowledge, skill, and ability or in the attitudes, values, and motivation they bring to their work. Since those leaving are no longer doing federal work, at least directly, it seems of less interest to attempt the assessment of whether their capabilities were, in the past, well matched to their job. We do, however, suggest that that be explored from the employee viewpoint as one of several opinion questions about perceived reasons for leaving government service.

Data Needed to Answer the Questions and Potential Sources

The question of capabilities lost when employees leave government is basically the same question posed in the previous chapter about capabilities to be measured in the core assessment. The data needs and the alternative sources that could be explored for information about capabilities of all kinds are the same. Two workforces are to be contrasted, differing only in that one group is still in government and the other has recently departed from government. The amount of experience being lost is of interest and for those leaving a federal job which is their first, educational indicators may also be of interest.

The sources for these needed data on education and work background are perhaps more limited in this part of the assessment than in the previous two parts. Those who have left by definition are no longer at work in federal positions and thus are not accessible on the job, so direct measures (tests, observations, etc.) would not be possible. A few data elements about promotion and award histories are available in the automated records. But the remainder of the needed information on workers, such as education, job history, or training, would be to some degree substantively incomplete, possibly out of date, and in any case inaccessible in paper records, as discussed in chapter 3, perhaps even more than usual since the separated employees' files may be segregated in storage.

College entrance test scores could be obtained by the matching methods already outlined, though they would provide interpretable data about only a fraction of those leaving.[1]

Employee (and former employee) self-reports, therefore, are once again the proposed source of the education, work history, and continuing education indicators. As discussed in both earlier chapters, supervisors are not a good source of information on employees' background or even their attitudes; more direct data are preferable. Employee self-reports are interpretable and are feasible to gather from current employees; locating separated employees (through the agencies) raises feasibility issues, but not impossible problems. How a sample of separated employees would weigh the positive and negative incentives when deciding whether to respond can only be tested in the early stages of the assessment. Response rates of other recent assessments of separated employees may not be generalizable since the surveys have been conducted on different populations such as the Senior Executive Service.

Questions about the views of those who left federal jobs cannot be answered by the attitude data from general surveys of current employees discussed in the previous chapter. The final personnel action on an employee who is leaving is to include a reason, but this piece of data is hard to interpret because the employee may not be truthful. The answer might affect the chances of obtaining future recommendations from supervisors and does constitute the agency's official answer to inquiries from unemployment insurance officials, which can affect benefits. Thus, some attitude questions should be asked on the same employee survey we are proposing be used to gather education and work experience data. Answers to these questions could give some understanding of positive and negative factors influencing the decision to leave. Further development work as the assessment matures can focus on exit interviewing or other data gathering approaches, so that employees will view the method as confidential and trustworthy and will therefore provide accurate information. These attitude data do not constitute a true causal study of why people left government; they are further indicators about the individuals, but a full exploration would require other data and measures to weigh a variety of possible explanations.

---

[1] College admissions test data would be available on part of the experienced workforce. Files at the testing firms go back about 23-29 years, so potentially there are scores on file for those under about age 38 in the case of the SAT, for example. The more significant limitation, however, is that test scores should be used as data only on those for whom few other sources exist, that is, those with little or no work experience. Since the distribution of professional and administrative employees leaving government by age and work experience is not known, we could not estimate the size of the group for whom test scores might be available and interpretable.

Since the match of employees' capability and the needs of the job is not to be explored extensively in this part of the assessment, it is not necessary to ask current or separated employees those detailed questions or to survey supervisors to obtain a second evaluation. It might be desirable to have some supervisory opinion data about separated employees, but there are several problems of feasibility and interpretability with a new survey of supervisors for the assessment. Simply locating former supervisors could be difficult, if some or all of the employee sample is drawn from a population leaving during several years, as the supervisors also could have changed jobs in that time. Supervisors' lack of clear memory is another difficulty. Coloration of the supervisor's report because of the circumstances of the person's leaving is a third.[2]

The 3 X 3 matrix in table 6.1 summarizes our evaluation of the alternative data sources.

Table 6.1: The Feasibility and Interpretability of Alternative Data Sources for Assessing Separated and Retained Employees

	Interpretability		
Feasibility	**Low**	**Medium**	**High**
Low	Employees' data in official records	Behavioral simulations evaluated by experts (impossible for former employees)	Direct tests of employees' knowledge, skill, and ability Direct observation of work (impossible for former employees)
Medium	College entrance exam scores Reason for separation shown on final official record	Former employees' self-reports of characteristics, attitudes	
High	Supervisors' opinion of employees' characteristics and attitudes	Employees' self-reports of characteristics and attitudes	

---

[2] A number of years of performance appraisal data are available now in automated records at OPM, but only on part of the workforce. In the single-digit form in which it is stored, this indicator is ambiguous and we do not propose to use it elsewhere in the plans for assessing workforce quality where we can obtain more interpretable data on supervisors' views of the employee's abilities and match to the job. However, if further analysis shows that at least three years of appraisals were available on the type of employee samples to be drawn for this segment of the assessment, which would permit some analysis of trends in the performance of those who leave, some might view those data as a useful addition.

Proposed Design

Occupations to Be Selected

As in the case of assessing quality in those entering the workforce, occupations should be selected for which possible losses of quality in the workforce are of particular interest. Unusual efforts by agencies in recruiting or retention may serve as a signal of occupations to be selected. Criteria for selection could emphasize:

- occupational series in which large numbers or proportions are quitting (4411 nurses left federal service in fiscal year 1987, for example, a rate of 12.1 percent, while 948 attorneys left, a rate of 6.7 percent);
- occupational series in which losses of quality staff may hurt performance of general services within government (such as computer specialists, of whom 904 quit in fiscal year 1987, or contract specialists, which lost 642);
- occupational series in which competition with the nonfederal sector is severe as shown by the size of the pay gap (such as high-level attorneys, engineers, and systems analysts, for which the 1986 pay comparability survey found the weighted median salaries of private sector workers were 30.7 per cent higher than those of federal GS-15 employees doing comparable work, $80,803 compared to $61,842).

Assessing a number of occupations at once allows for efficiencies at all stages of the work, though there is nothing about measurement or analysis that dictates how many occupations are assessed or when. For estimating costs, we have assumed that 12 occupations would be covered, all surveyed at once.

Sampling

The CPDF transaction files provide a source of all those leaving government. We suggest examining those leaving voluntarily; very few leave involuntarily. Within each occupation selected for study, a group of those who left can be sampled and the agency they left can be requested to provide the last known address. For mailing tax and other information, the address is usually maintained for at least a year. Further technical development can explore the question of the exact definition of separation to be used in forming the group to sample. In addition to sampling those voluntarily leaving other than by retirement, there may be value in selecting also from a group who appear to be retiring at an unusually early age. This discussion assumes that the full range of years of service is examined. It would be possible to restrict the study to only those with substantial government service, if it were believed that quality issues were most salient there. Loss of the most capable younger employees is of course also possible and quality indicators should be selected to provide interpretable data on all levels of experience included in any sample.

The CPDF can be used to draw the comparison group samples, which would also be surveyed. These should be from continuing employees in the selected occupations, matched further in age, sex, and years of government service, so that, as far as possible, demographically similar groups are being compared.

Indicators

The indicators of capability would be similar to those sought in the core assessment including information on education, especially later augmentations through training; licenses or certifications, especially if some fields have these at more advanced levels, indicating specializations or updated skills (such as re-certifications); and work experience including jobs, promotions, and awards. The list should be shortened as specific samples are determined; for example, grade-point average would be reliably recalled only by those recently in school and would be interpretable as an indicator of quality primarily for those with little work experience. This item could be omitted if it is decided to focus on experienced employees only.

In the area of attitudes, values, and motivation there is room for considerable development. Unlike the core assessment in which the goal is a general evaluation of quality and attitudes are of unknown relevance, with separated employees there is a specific event whose causes and consequences can be asked of the former employee. Thus a set of questions can be asked that would explore the separated employees' work experience and values, what led to the decision to leave, and, if the person has a new job, what it is and what the individual finds more attractive in the new situation. The answers will not be completely reliable, owing to the natural inclination to attribute favorable results to one's decisions, but some useful generalizations can result about who is going where and what reasons appear to have motivated the moves. One such opinion would be roughly parallel to the job-match concept studied more directly in the core assessment. Separated employees could be asked if their knowledge, skill, and ability are being used to a greater or lesser extent in the new job. Once attitude questions are designed for the separated group, similiar ones can be asked in any general employee attitude survey (described in chapter 5) to provide direct comparisons to the overall workforce. Table 6.2 summarizes the indicators we believe are useful.

Table 6.2: Dimensions, Indicators, and Data Sources for Assessing Separated and Retained Employees

Dimension and indicators	Source of data
Knowledge, skill, ability	
Education	
Years of schooling	Employee survey
Degree(s) awarded	" "
Date of degree(s)	" "
Grade-point average	" "
Rank in class	" "
Institution attended	" "
Major field of study	" "
Continuing education, training	
Quantity (hours, days, units, credits)	" "
Kind (course names)	" "
Source	" "
Professional certificates, licenses	
Examination scores (CPA, bar)	" "
Certification, licensure record	" "
Work experience	
General work experience	" "
Specialized work experience	" "
Promotion history	" "
Awards (monetary and other)	" "
Attitudes, values, and motivation	Employee survey (separated employees only)
Match of individual capacities and job needs	Employee survey (separated employees only)

Summary and Conclusions

Useful data can be obtained on the quality of those leaving government jobs. Quality would be defined in terms of employees' capabilities, including their knowledge, skill, and ability, and their attitudes and values toward work. Data on those leaving would come from surveys of samples in selected occupations and would be compared to similar data obtained from employees still in the same federal occupations.

Data gathered according to the proposed design would help provide generalized answers to questions such as:

- When top scientists leave government, do they tend to continue doing government-related work, although for a new employer, or are they lost to federal missions?
- What is the profile of the quality of attorneys leaving government, in terms of their legal education as well as their career paths once in government? Where are they going and what do they say about their reasons?
- Are the nurses leaving government service more educated on average than the ones who stay?
- What is the average length of service of the most qualified people leaving government in these occupations — that is, do highly capable people more commonly leave at early or late points in their federal careers?

Strategies for Implementing the Assessment

The assessment design described in chapters 4-6 is summarized in table 7.1, which shows the three segments of the workforce proposed for evaluation and the relevant questions, data sources, and analyses that we have suggested for each. (See table 7.1.)

Ensuring the adequate implementation of this design is as important as planning the study details. The appropriate structure is important because the task is complex, includes conceptual, methodological, and data problems, and will involve some continuing federal funds, though not a large amount. Questions about implementation of the proposed assessment addressed in this chapter include:

- What criteria should be used in considering the organizations within or outside the federal government that could be responsible for the assessment?
- What funding issues should be considered as part of planning for the needed work?
- Are there preferable approaches to timing and phasing segments of the assessment, if it cannot all be started at once?

We conclude that a data base is feasible that can help answer the questions raised in the committee's request, and we recommend that the Congress should establish the assessment by direction to the executive branch.

Table 7.1: Design for an Assessment of Federal Workforce Quality

Segment of workforce and pertinent questions	Indicators	Data sources
Entering workforce How good are those attracted to federal jobs and how do they compare with others hired elsewhere?	Knowledge, skill, and ability (as indicated by education and work experience)	Employee self-reports, test score records
Current workforce How good is the workforce and how has that changed over time?	Knowledge, skill, and ability; attitudes, values, and motivation, match of capabilities to job needs	Employee self-reports, existing attitude surveys (with modification), special supervisor appraisals
Separations from the workforce Is the federal government losing its best employees?	Knowledge, skill, and ability; attitudes, values, and motivation	Employee self-reports

Organizational Arrangements

Diverse tasks are involved, including tasks of policy (deciding on the scope and objectives of the assessment), research and development (deciding on measurement instruments, sampling, pilot testing, refinement of measures in later cycles), logistics (contracting for survey support, handling data tapes), analysis (interpreting the data), communication (reporting the results), and management and coordination.

In addition, the assessment program should be located in a setting in which it can be seen not as a one-time project, but as a continuing activity. This perspective may help to insure that trend analyses can be based on consistent measures and methods of interpretation that are now missing from the scattered discussions of workforce quality. A long-term perspective is also needed so that there will be support and direction for such additional activities as:

- doing research and development to refine the techniques used (as suggested in many places in the earlier chapters);
- reaching out to stimulate interest in agencies, policy research groups, and universities in strategic assessments of the workforce (comparing existing and needed quality); and

Comparison	Estimated sample sizes	Frequency	Analysis
Entering nonfederal employees in similar occupations	2,400 new federal employees, 2,400 nonfederal (in 12 occupations)	As desired	Within occupations, at one point in time, comparison of federal and nonfederal new hires on education and experience
Federal employees in the same occupations at earlier times, supervisor and employee perceptions at the same time	27,600 employees and their supervisors (in 23 occupations)	4-year intervals	Within occupations, across several points in time, trends in education and experience, attitudes, and degree of perceived match of employee capabilties to individual jobs
Nonseparated federal employees in the same occupations	3,600 separated employees, 3,600 matched nonseparated (in 12 occupations)	As desired	Within occupations, at one point in time, comparison of those staying and those leaving on education, experience, and views

- responding to those wishing to use the data including preparing public data tapes, or perhaps sponsoring secondary analyses that can extend the interpretation of the rich data that will accumulate.

Criteria that should be considered in analyzing where to locate the administration of the assessment include:

- existence of the required expertise and experience with the methods and data required for assessment;
- access to data bases needed for sampling, comparison studies, etc.;
- objectivity with respect to the outcomes of the assessment;
- commitment to the effort and its stability over time (e.g. personnel, funding); and
- availability of resources to carry out the tasks.

We believe that such a continuing program of assessment of workforce characteristics properly belongs in the executive branch. Within the executive branch, OPM has responsibility for workforce data, including the official personnel folders held at the agencies. OPM maintains the most comprehensive automated workforce data base, the CPDF, which can be considered a subset of the folders' data. OPM also has responsibility for assisting agencies with obtaining the quality of staffing needed (including developing tests and other examination methods to determine quality) and for evaluating personnel management. This evaluation could include using, among other criteria, measures of the quality of workforce attracted and retained. OPM is required to make several reports to the Congress on the workforce. As the agency with the most detailed knowledge of both the workforce and methods for assessing individual qualities, OPM is the logical candidate for performing the proposed assessment.

Nevertheless, OPM must be assessed on the other criteria as well. OPM might not be able to be fully objective with respect to the outcome of the assessment, since quality problems reflected in the assessment data could reflect on staffing practices in which OPM was involved or on other policies on which OPM had advised officials of the executive branch. Further, OPM lost research capacity in the early 1980s. OPM thus may not have adequate resources to carry out the tasks and may not be able to make the necessary commitment to a stable design and repeated assessments, though this could change in the future.

Relevant expertise in examining labor market issues generally, including sampling and surveying employers, is available at the Bureau of Labor

Statistics. The bureau has dealt with issues of independence, timeliness, and technical quality in gathering and reporting economic indicators that are of intense public interest. The subject matter of the workforce quality assessment may be more comprehensive (in the types of indicators involved) than that of many projects in the bureau, however, and there are no current resources available for the task.

Another possible organizational location for the assessment is the Merit Systems Protection Board, an independent agency created by the Civil Service Reform Act of 1978 to oversee and report on all aspects of the laws governing the federal workforce. In connection with its oversight of the personnel system and to provide data for a mandated report on significant actions of OPM, the board has carried out a number of research and evaluation projects concerning the federal workforce. The board by law is to be bi-partisan in composition. The board staff has included experts in personnel studies and survey research in recent years, though fluctuating in extent, and there are no current resources for additional tasks. Specific direction to take on the new work and assignment of specific resources would be needed to insure that the assessment would be a stable part of the board's research and evaluation portfolio.

Private organizations with experience in personnel research or general surveys are probably not suited to the multiple tasks involved in the overall assessment, though appropriate contracts for specific tasks could be used. Such organizations would lack ready access to federal agency records and data bases and would have weak influence over agencies' responsiveness to the needs of the assessment.

An advisory board of individuals with experience and expertise in gathering and using data on workforces could be helpful to the organization eventually selected to perform the assessment. The group could include experts from various disciplines of data-gathering and analysis, as well as experienced personnel officials and managers from public and private sectors, along with individuals representing eventual users of the data from Congress, the press, and employee groups.

An implementation structure which should be considered would be to designate one agency as the responsible organization to carry out the assessment, but to mandate others as well to provide specific services, including BLS for help with comparison groups of nonfederal workers, OPM for drawing statistical samples using the CPDF, MSPB for the employee attitude data, and any others that seem useful (Bureau of the

Census, Department of Education, etc.). Specific oversight roles for an external advisory board should also be designated from the outset, including review of designs, instruments, sampling and analysis plans, the resulting data and interpretive products.

Funding for the Assessment

Although there are limits to what can be known about the quality of the federal workforce at the present time, with adequate funding and personnel a responsible organization could do a good job in tracking indicators of quality for different segments of the professional and administrative group. If the assessment is to be done, however, new resources are needed, since at present no organization has a usable data base (as detailed in chapter 3) and no organization has such a task in its mandate or current plans.

If the proposed assessment is established, specific resources should be authorized for a substantial term. The results of the assessment will be more useful if the study framework is stable, as that helps insure comparable data across time, and such stability is uncertain without special precautions. If the government-wide employee attitude survey data are considered important, specific direction to MSPB may be needed as well, as there is no assurance that the board will be able to continue the survey at regular intervals or will maintain stable methodologies which encourage valid comparisons (similar samples, question wording, etc.).

So that the data from the assessment can be used by various analysts, additional funds could be useful for a program of independent sponsored research using the database. Staff and resources could also be made available for synthesizing studies done by individual agencies that could provide benchmarks or comparisons to the assessment data or for smaller studies in a few agencies or specific occupations in which targets of opportunity arise. For example, once the instrument package is ready, the assessment instruments could be used to evaluate effects on the quality of the workforce resulting from personnel policy demonstration projects such as the Navy's China Lake experiment. In this way, the assessment group could represent a body of knowledge and expertise to expand the effort to assess workforce quality beyond the assessment itself. Funds should in any case be provided to prepare documentation for public use computer tapes of the assessment data, which can be deposited in the National Archives' collection of machine-readable data.

A detailed cost estimate for the assessment is included in appendix III, which shows that initial assessments of entering and exiting employees in selected occupations (together with necessary comparison groups), plus collecting the baseline data for the core assessment trend analyses, would involve an estimated direct federal cost of approximately $718,000. Some costs cannot be accurately estimated at this point and in particular the key comparison study of new-hires in nonfederal organizations will involve unknown but possibly substantial costs. The total costs, both new expenditures and the cost of employee time to complete survey instruments, appear modest, however, in comparison to the potential value of usable data on the quality of results obtained from the substantial time and funds spent in recruitment, selection, and other efforts to attract and retain a high quality federal workforce.

Timing and Phasing of the Assessment

Because of the widespread discussion of the quality of those in public service, the assessment should begin as soon as possible. In anticipation of the transition to a new administration in 1989, followed by possible changes in program policy and personnel policy, it would be important to gather baseline data quickly on the quality of those in the federal workforce now, before such changes take effect.

Table 7.2 shows the relative complexity and difficulty of the three parts of the assessment, as outlined in chapters 4-6. Each poses distinctive challenges, including obtaining the comparison group of nonfederal new hires for the first segment, efficiently managing the overall scale of the stratified samples and multiple instruments to be used in the core assessment, and locating the former federal employees and obtaining their cooperation in providing data for the third segment. (See table 7.2.)

In terms of scope, an effort such as this could be phased in or all components initiated concurrently. Each approach has both benefits and risks. One could, for example, begin the assessment on a more modest scale by gathering less data, such as by using only a few indicators of quality. However, those indicators that are most accessible, for instance, the college entrance test scores stored in computer files at the testing organizations, are limited in interpretability; publicizing results on any isolated measure, according to many of the people we consulted in doing the review, could damage the eventual acceptability of the overall assessment. Another approach would be to begin either the new employee or separated employee studies without comparison groups (to reduce the data-gathering load). We see great risks to interpretability if corners were cut by eliminating the comparison groups. To omit them invites

Table 7.2: Relative Difficulty of Parts of the Assessment

Segment	Sample	Degree of Difficulty
Entering workforce	Federal	Not difficult
	Nonfederal (comparison group)	Difficult (establishing universe, locating sampled respondents)
Core assessment	Overall	Moderately difficult because of overall complexity of assessment dimensions and double sample (employees and supervisors)
Separating workforce	Nonfederal	Moderately difficult (establishing universe, locating sampled respondents, obtaining cooperation)
	Federal (comparison group)	Not difficult

more of the same speculation (for example, on whether other employers are more successful in attracting quality in competition with the federal sector), which is precisely what the overall assessment is designed to reduce.

One way to start smaller might be to begin the core assessment, which includes the most comprehensive definition of quality and the most interpretable set of indicators but to reduce the workload by starting with fewer selected occupations than the full set of 23 listed in appendix II. Those occupations not selected for the initial year's work could be assessed for their first time in subsequent years. Trend analyses would thus be done on different groups in different years, for example, attorneys in the years 1988 and, if the interval is four years, 1992, while nurses could be studied in 1989 and 1993. Each initial occupational sample would include some employees in the first years of service, which would provide the limited picture of quality of those entering (without the important comparison group, as just described), though these results would be able to be compared with the fuller picture on the rest of the workforce gathered at the same time. Conducting the assessment in this way, however, would not allow the major current questions to be answered and can only be recommended as a contingent measure.

Conclusions

GAO believes a data-collection program such as that outlined in chapters 4-6, established at congressional direction and with sufficient resources to carry out the work in a methodologically sound way, would provide useful information. The results could satisfy the strong congressional interest in the quality of the federal workforce and could also aid in decisions on personnel policy in the executive branch.

An assessment of federal workforce quality employing the definition proposed in chapter 2 is feasible, using new data from samples of federal employees and others. Three major questions would organize the inquiry, corresponding to three parts of a model of the workforce as composed of a core — the current workforce—with flows in and out. The core assessment would determine changes in the overall quality of the workforce over time. The other two assessments would answer questions about the quality of those coming into and leaving federal jobs based on specific comparisons to other relevant groups.

Because federal agency data are now not designed for the purpose and because no agency has a mandate to do such an assessment, a good deal of new work is necessary, which will take time to produce the full anticipated results. Therefore the organizational and funding arrangements are important to consider so that the assessment can count on the needed expertise, commitment to the task, independence, and stability.

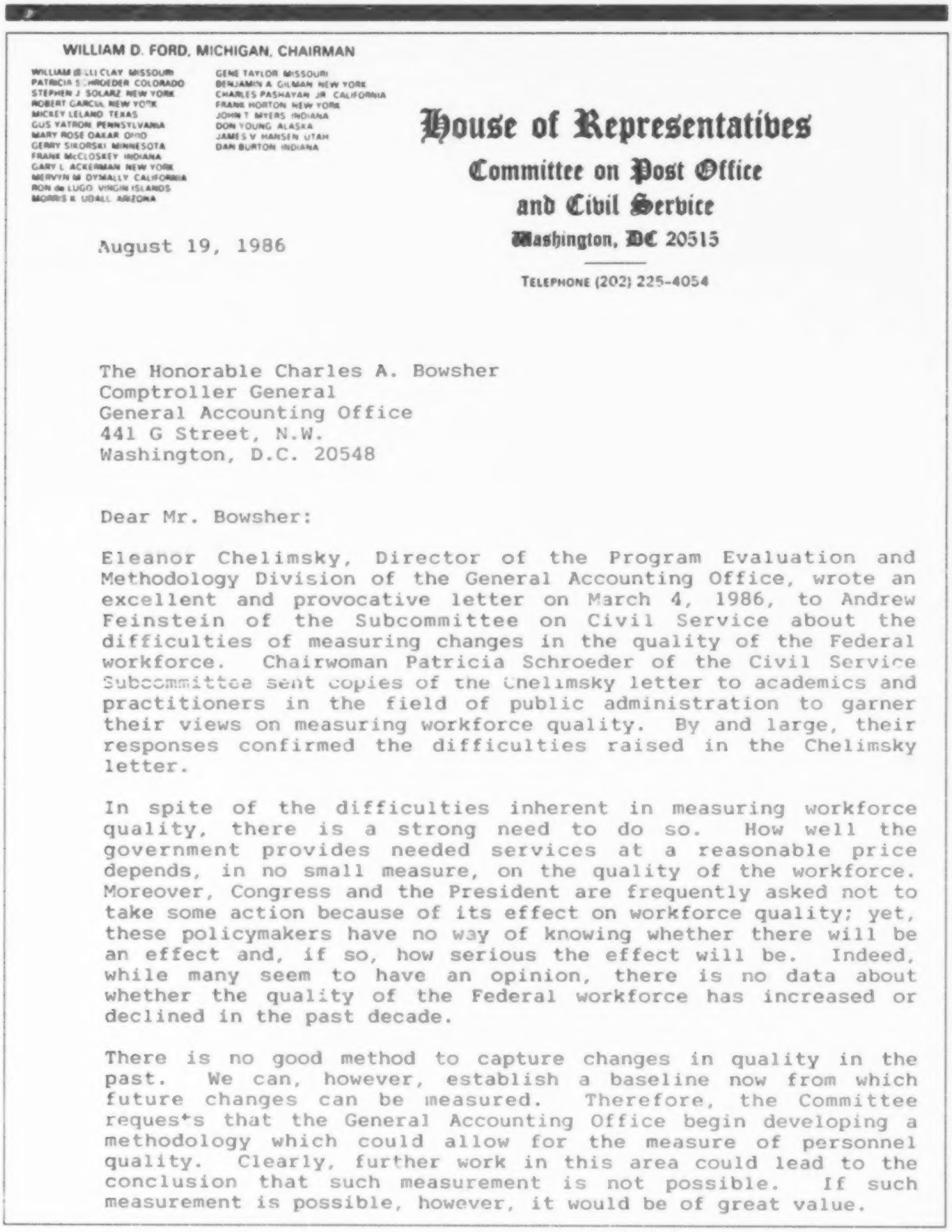
Recommendation to the Congress

To help provide generalizable answers to questions about the quality of the federal workforce and changes in it over time, we recommend that the Congress authorize a continuing program to assess workforce quality based on the framework outlined in this report.

Agency Comments

We discussed the proposed assessment with officials at the Office of Personnel Management and the Merit Systems Protection Board. The officials in general raised no major problems with the proposed definition of quality, the initial indicators, or the broad feasibility of the designs for the three segments of the assessment. However, in accordance with the requester's wishes, we did not obtain written comments on a draft of this report.

Congressional Request Letter



WILLIAM D. FORD, MICHIGAN, CHAIRMAN

WILLIAM (BILL) CLAY MISSOURI
PATRICIA SCHROEDER COLORADO
STEPHEN J. SOLARZ NEW YORK
ROBERT GARCIA NEW YORK
MICKEY LELAND TEXAS
GUS YATRON PENNSYLVANIA
MARY ROSE OAKAR OHIO
GERRY SIKORSKI MINNESOTA
FRANK McCLOSKEY INDIANA
GARY L. ACKERMAN NEW YORK
MERVYN M. DYMALLY CALIFORNIA
RON de LUGO VIRGIN ISLANDS
MORRIS K. UDALL ARIZONA

GENE TAYLOR MISSOURI
BENJAMIN A. GILMAN NEW YORK
CHARLES PASHAYAN JR. CALIFORNIA
FRANK HORTON NEW YORK
JOHN T. MYERS INDIANA
DON YOUNG ALASKA
JAMES V. HANSEN UTAH
DAN BURTON INDIANA

House of Representatives
Committee on Post Office
and Civil Service
Washington, DC 20515

TELEPHONE (202) 225-4054

August 19, 1986

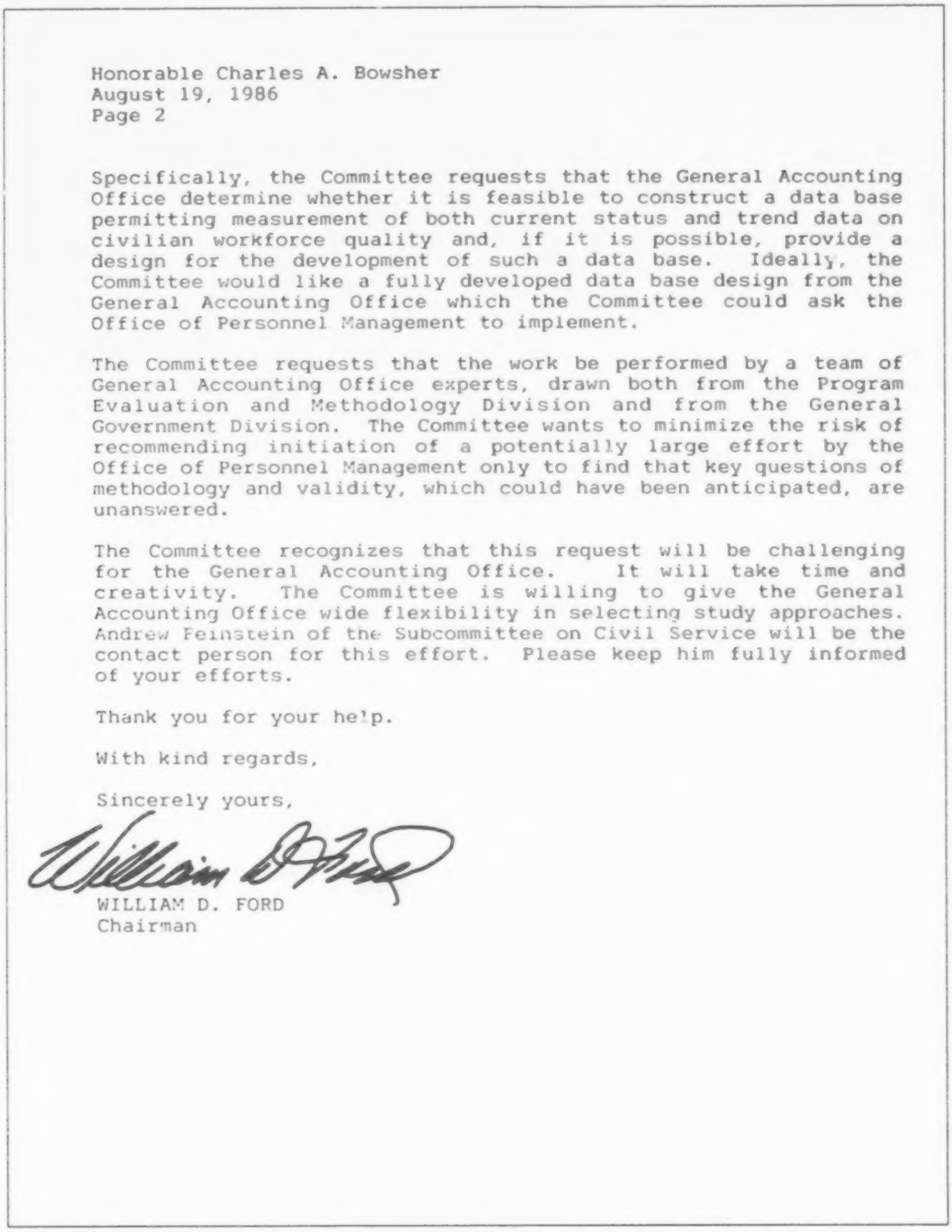
The Honorable Charles A. Bowsher
Comptroller General
General Accounting Office
441 G Street, N.W.
Washington, D.C. 20548

Dear Mr. Bowsher:

Eleanor Chelimsky, Director of the Program Evaluation and Methodology Division of the General Accounting Office, wrote an excellent and provocative letter on March 4, 1986, to Andrew Feinstein of the Subcommittee on Civil Service about the difficulties of measuring changes in the quality of the Federal workforce. Chairwoman Patricia Schroeder of the Civil Service Subcommittee sent copies of the Chelimsky letter to academics and practitioners in the field of public administration to garner their views on measuring workforce quality. By and large, their responses confirmed the difficulties raised in the Chelimsky letter.

In spite of the difficulties inherent in measuring workforce quality, there is a strong need to do so. How well the government provides needed services at a reasonable price depends, in no small measure, on the quality of the workforce. Moreover, Congress and the President are frequently asked not to take some action because of its effect on workforce quality; yet, these policymakers have no way of knowing whether there will be an effect and, if so, how serious the effect will be. Indeed, while many seem to have an opinion, there is no data about whether the quality of the Federal workforce has increased or declined in the past decade.

There is no good method to capture changes in quality in the past. We can, however, establish a baseline now from which future changes can be measured. Therefore, the Committee requests that the General Accounting Office begin developing a methodology which could allow for the measure of personnel quality. Clearly, further work in this area could lead to the conclusion that such measurement is not possible. If such measurement is possible, however, it would be of great value.



Honorable Charles A. Bowsher
August 19, 1986
Page 2

Specifically, the Committee requests that the General Accounting Office determine whether it is feasible to construct a data base permitting measurement of both current status and trend data on civilian workforce quality and, if it is possible, provide a design for the development of such a data base. Ideally, the Committee would like a fully developed data base design from the General Accounting Office which the Committee could ask the Office of Personnel Management to implement.

The Committee requests that the work be performed by a team of General Accounting Office experts, drawn both from the Program Evaluation and Methodology Division and from the General Government Division. The Committee wants to minimize the risk of recommending initiation of a potentially large effort by the Office of Personnel Management only to find that key questions of methodology and validity, which could have been anticipated, are unanswered.

The Committee recognizes that this request will be challenging for the General Accounting Office. It will take time and creativity. The Committee is willing to give the General Accounting Office wide flexibility in selecting study approaches. Andrew Feinstein of the Subcommittee on Civil Service will be the contact person for this effort. Please keep him fully informed of your efforts.

Thank you for your help.

With kind regards,

Sincerely yours,

William D Ford

WILLIAM D. FORD
Chairman

Selection of Occupations for Sampling

The core assessment described in chapter 5 should examine indicators of quality in the workforce within a consistent set of occupations across time. (Of course the set could change over the long term, after at least two rounds of data-collection establish the general trends in indicators for any occupation.) This appendix describes an approach to selecting the occupations from which employees would be sampled, for that part of the assessment only, based on the size of the workforce in each occupation.

Using OPM data for 1985 (the latest year for which statistics on occupations are available), we selected all major professional and administrative occupations paid on the general schedule or its equivalent. We defined "major" as any occupation with at least 8,000 employees. (We eliminated the 1710 series, occupational and vocational training, because the majority of employees are not on the general schedule or an equivalent pay system, and the 861 series, aerospace engineer, as the list already included 4 other kinds of engineers.)

In addition, we suggest forming two occupational clusters, one of 13 physical science jobs (such as physicist, chemist, and geologist) and another of 12 biological science jobs (such as ecologist, botanist, and zoologist). Each of the individual occupational series has fewer than 8,000 employees, but as the combined groups with similar qualifications and position descriptions are sizable (32,508 in the physical science cluster and 9,878 in the biological science cluster), and as there is continuing concern for the quality of scientific personnel in government, we believe it is useful to include them.

These occupations appear to be an adequately comprehensive and balanced set from which to sample. The total of 435,000 employees includes about 53 percent of the overall professional and administrative workforce. Of the set of 23 occupations (considering the scientific clusters as two occupations) we are proposing, 12 are professional and 11 are administrative. The overall proportion of women is 26 percent in professional and 34 percent in administrative occupations; this group of occupations includes 31 percent women.

Table II.1 shows the 23 occupations meeting our criteria. Table II.2 shows the two sets of scientific occupations that have been combined.

Table II.1: Occupations Selected for Quality Assessment

Series	Title	Type[a]	Total employees	Men	Women
105	Social Insurance Administrator	A	21,401	9,626	11,775
201	Personnel Management Specialist (includes Personnel Officer)	A	9,471	4,827	4,644
334	Computer Specialist	A	40,122	27,493	12,629
341	Administrative Officer	A	8,460	3,805	4,655
343	Management Analyst	A	15,694	8,164	7,530
345	Program Analyst	A	16,162	8,850	7,312
510	Accountant	P	11,198	8,006	3,192
511	Auditor	P	12,435	10,079	2,356
512	IRS Agent	P	14,293	10,623	3,670
560	Budget Analyst	A	11,359	4,522	6,837
602	Medical Officer[b]	P	9,600	7,864	1,736
610	Nurse[b]	P	39,109	3,419	35,690
801	General Engineer	P	19,569	18,863	706
810	Civil Engineer	P	16,775	15,847	928
830	Mechanical Engineer	P	13,583	12,927	656
855	Electronics Engineer	P	24,033	22,952	1,081
905	General Attorney	P	17,796	12,962	4,834
1102	Contracting Specialist	A	27,871	14,246	13,625
1811	Criminal Investigator	A	24,230	22,303	1,927
1910	Quality Assurance Specialist	A	16,829	14,946	1,883
2152	Air Traffic Controller	A	22,660	20,454	2,206
Sub-total, individual series			**392,650**	**262,778**	**129,872**
	Biological Researcher	P	9,878	7,748	2,130
	Physical Scientist	P	32,508	28,196	4,322
Total, all series			**435,036**	**298,712**	**136,324**

Source: Office of Personnel Management, Federal Civilian Workforce Statistics: Occupations of Federal White Collar and Blue Collar Workers (Washington, D.C.: October 31, 1985).

[a] Indicates the type of job, showing for each series the OPM classification as either professional (P) or administrative (A). Professional series have positive educational requirements for hiring, while administrative series do not.

[b] These occupations are paid on a pay schedule equivalent to the general schedule.

Table II.2: Additional Occupations to Be Grouped for Quality Assessment

Series	Title	Total employees	Men	Women
	Biological Researcher			
401	General Biological Scientist	4,752	3,778	974
403	Microbiology	1,848	1,166	682
405	Pharmacology	439	338	101
408	Ecology	325	272	53
410	Zoology	120	99	21
413	Physiology	451	367	84
414	Entomology	745	706	39
430	Botany	173	119	54
434	Plant Pathology	319	287	32
435	Plant Physiology	307	273	34
437	Horticulture	117	96	21
440	Genetics	282	247	35
Total		**9,878**	**7,748**	**2,130**
	Physical Scientist			
1301	General Physical Science	5,323	4,742	581
1306	Health Physics	599	518	81
1310	Physics	4,278	4,073	205
1313	Geophysics	607	552	55
1315	Hydrology	2,249	2,049	200
1320	Chemistry	7,602	5,992	1,610
1321	Metallurgy	444	426	18
1330	Astronomy and Space Science	543	505	38
1340	Meteorology	2,133	2,017	116
1350	Geology	2,601	2,223	378
1360	Oceanography	761	679	82
1370	Cartography	4,985	4,077	908
1372	Geodesy	383	333	50
Total		**32,508**	**28,186**	**4,322**

Source: OPM (same as table II.1). Data are as of October 31, 1985.

Appendix III

Sample Sizes and Costs

Sample Sizes

To project costs for the proposed assessments, sample size is a major determinant. When precise characteristics of measuring instruments are not yet determined, as in the case of the proposed employee and supervisor surveys to be used in the three segments of the design described in chapters 4-6, estimates of sample size must rely on assumptions. The other major parameter affecting sample sizes is the degree of confidence in the results that is desired. The greater the confidence or precision desired, the larger the sample must be.

To estimate the initial sample sizes, we have used conservative assumptions about the characteristics of the instruments and the most stringent degree of confidence (that the sample estimate be wrong in either direction by no more than 5 percentage points). These planning assumptions result in a needed sample of 400 for each group on which conclusions are desired.[1]

Sampling the Entering Workforce

The assessment of the entering workforce described in chapter 4 involves only one level of experience (all those surveyed are newcomers) and we propose that fewer occupations be studied than the 23 nominated for the core assessment. Thus the needed samples are relatively small. The 400 new hires needed for analysis of an occupation should be split between federal and nonfederal employees.[2]

The total sample size for this segment depends on the number of occupations to be examined. Assuming that interest in new hires is not as intense for all 23 of the occupations in the core assessment, we project a total group of occupations half as large or 12 total. Assuming also that analysis will be limited to the examination of differences, 12 occupations x 400 gives a sample of 4,800. Half would be federal employees and half nonfederal.

Sampling the Current Workforce

The overall professional and administrative workforce should be assessed using a stable and relatively comprehensive set of occupations, as outlined in chapter 5 and appendix II. Sampling 400 employees at each of three levels of experience within each of 23 occupations will result in an employee survey sample totalling 27,600. Each sampled employee's supervisor is also surveyed. Not all employees and supervisors would be surveyed at one time, however, if the assessment plans assign different jobs to different four-year intervals.

---

[1] The effect of relaxing confidence requirements is dramatic. If the tolerable margin of error is 8 percentage points the required sample drops to 160; with an acceptable margin of 10 percent, the sample can be 100.

[2] This allows acceptable estimates of differences between federal and non-federal employees in each occupation. If it was desirable to make separate estimates on the survey variables for each group, then the sample size would again be 400 per group per occupation, effectively doubling the sample size.

Sampling the Exiting Workforce

For studying separated employees, a sample of 300 (matched to 300 who stayed in government), or a total of 600, is needed per occupation. The number of sampling units, 300 pairs of employees in this case, is smaller than in the previous discussion where 400 units (individuals) are needed in an occupation for the most careful estimates. Fewer sampling units are needed since each separated employee is matched with another nonseparated employee on several dimensions, and the overall variance will therefore be somewhat less. With smaller expected variance the number of units sampled can be reduced and still attain the same effective degree of precision. The pairs will be matched on age, sex, and years of government experience, so formal stratification by experience is not needed in the sampling. Total sample size again depends on the number of occupations to be examined. The quality of people leaving is critical in some jobs but probably not in all of the 23 suggested for the core assessment, so we have once again estimated the total sample size using 12 occupations. No supervisor data are to be gathered on this group. For 12 occupations x 600, the sample size would be 7,200.

Sampling to Obtain Agency-Level Comparisons

Pinpointing quality differences by agency in any of the three segments of the assessment would require adding another stratification scheme to the sampling design. It is not likely that there would be enough people in a range of agencies in the basic samples for statistically sound comparisons. The planning would need to consider the fact that some jobs are highly concentrated in a few agencies, while others are dispersed. For jobs in which a large number of the workforce are employed at a few agencies, it would be especially feasible to sample employees in each job by agency with confidence that adequate numbers are there for analysis. For example, for 11 of the 23 jobs we propose for assessment (shown in table II.1 in appendix II), 66 percent or more of the workforce in each of the jobs can be found in three or fewer agencies. (For six of the 11 job series, the concentrations are in different sets of agencies. The other five include three for which the three military services account for 66 percent or more of the workforce, and medical officers and nurses which are both concentrated at the VA, HHS, and Army.) Meaningful agency comparisons would appear to be possible in those cases. For jobs that are more dispersed across numerous agencies, and for which there may be smaller numbers at each agency, acceptable agency-level samples may not be possible. There appear to be some special problems in attempting agency-level analysis in the study of entering employees, also.[1] Further technical development should consider the trade-offs among feasibility, cost, and interpretability for various sampling approaches.

Table III.1 summarizes the initial samples for each segment of the assessment, without provision for agency-level analyses.

Table III.1: Sample Sizes Needed for Three Segments of the Proposed Assessment of Federal Workforce Quality

Assessment segment	Number of occupations	Other strata	Total group
Entering employees			
Federal employees	12[a]	None	2 400
Nonfederal employees	12	None	2 400
Current employees			
Employees	23[b]	3	27 600
Supervisors		None	27 600
Exiting employees			
Separated employees	12[a]	None	3 600
Current employees	12[d]	None	3 600

[a] Number of occupations for estimating purposes only actual number could be larger or smaller depending on final design of the assessment

[b] Recommended number of occupations see appendix II

[c] Supervisors would not be sampled but would be identified as the supervisors of record for the sampled employees

[d] Within each job series selected for study current employees would be chosen by matching those leaving on sex age and experience

Cost Estimate

The two basic cost elements of the proposed assessment program are the staff time for development and analysis and the contract costs of survey administration and data preparation. Large-scale surveys are sometimes done by government employees (such as those in the Census Bureau) but are more commonly contracted to specialized private-sector firms. The core assessment we estimate would involve up to four staff years of

---

[1] Relatively small numbers may be hired in a single occupation in specific agencies in one year: solving this problem by drawing a survey sample from an aggregated group of several years' new hires raises other problems. For example, employees sampled from CPDF records showing them to be working in a particular agency may, by the time they actually complete a survey instrument, have changed agency. And if the span for agency-level sampling has to be as broad as three years to find needed numbers of new people in a job series, in that period there may have been too many changes in policy and work conditions to permit the desired clear-cut analyses of effects on the quality of people attracted.

time for development and pretesting and for analysis after the data are received in usable form. The entry and exit segments of the design would not add appreciably to the staff effort needed in instrument development, but as they present some unique sampling and analysis challenges, we have added another half staff-year. For per-unit survey processing costs, we have used the average cost experienced by MSPB in a recent contract for processing its Merit Principles Survey, which went to about 21,000 federal employees in a manner similar to that we are proposing. MSPB officials estimated that they paid a contractor $1.50 per survey for printing and distributing the surveys, managing their receipt, entering the data into computers (using survey response forms that can be read automatically by optical scanning equipment), and producing a usable data file for analysis.

Costs of drawing various samples of federal employees are simply the computer and staff costs for the OPM data center, as the samples can all be drawn from OPM's Central Personnel Data File. This is the case for all samples in the proposed program except the comparison group of nonfederal new hires, which presents special problems.

Costs of Alternatives for Sampling the Entering Workforce

As discussed in chapter 4, there are significantly different options for locating the necessary comparison group of nonfederal new hires for that segment of the assessment. In one method, people new to their jobs could be found by asking a sample of employers to identify those they have hired in specific occupations within a specified time period. In a second method, a group of recent college graduates, who by definition will be new to any job, could be traced starting with samples from school records.

The approach through employers has the advantage of greater coverage of the new hire group, including both inexperienced and experienced individuals new to their jobs. It has the disadvantage of uncertain but potentially significant costs and burdens, as the state and federal agencies who maintain employer lists do not have experience surveying employers (or working through them) for the particular purpose of obtaining information on individuals. The Bureau of Labor Statistics reported on their experience gathering routine data series on employment and the approach the states use to reach samples of firms. Bureau officials said they have also used the same approach for one-time studies of national populations of employers, such as a recent survey on employee drug-use. But all the data gathered using the employer sampling frames are summary statistics on groups of people. Because the costs of many activities in the joint BLS-state statistical reporting on occupations are included in cost-sharing agreements negotiated with the states, BLS could not estimate the specific costs of adding our particular requirements to existing agreements.[1] BLS officials did not recommend working through states, but concentrated on options based on drawing national samples of firms likely to have new hires in the occupations of interest, using BLS merged files of employers drawn from state records. Even so, because the process would involve several steps beyond any they have attempted with employers, they could not be sure that the survey effort would be feasible or what it would cost. We explored three alternatives: employers could supply information about individuals' backgrounds, drawn from files, though we doubt this data source would be comprehensive, as discussed in chapter 4; employers could distribute surveys to employees to be returned to the assessment (an approach similar to that to be used with federal employees); or employers could provide addresses which the assessment agency or BLS could use to mail surveys. If an employer had hired large numbers of people in the target occupations, there would need to be a method for drawing a sub-sample within that group, to lessen burden.

The feasibility of the second approach, finding new hires by tracing recent college graduates, is shown by surveys done by the Department of Education. The department periodically surveys a national sample of people who have just received bachelors' and masters' degrees, about one year after their degree date. The survey does locate some who have taken federal jobs (572 federal employees responded to the last department survey, along with about 10,000 others), and thus appeared to include the two groups needed for this assessment within its design. But we found that the group of federal employees is too diverse and the numbers of people in any particular federal occupation too small to permit analysis of quality indicators within specific professional and administrative occupational series, which is our design requirement. Attempting to augment the sample to insure enough federal employees

---

[1] As an example, BLS officials told us that a federal agency that needs detailed data on scientists and engineers transfers $240,000 per year to augment data-gathering on those occupations alone.

(about 200 are needed in each selected occupation) raises questions of feasibility and cost.[illegible]

Thus the greatest potential use of the department's survey would be to provide the comparison group data only. That is, entering federal employees could be surveyed directly using a sample drawn from OPM files, while data on those in some similar nonfederal jobs could be drawn from those sampled in the Department of Education survey. (Chapter 4 discusses limitations of this comparison group.) We reviewed the department's data from the most recent survey to see if there were the 200 nonfederal employees needed for analysis in the 23 occupational series we selected. Department officials told us only three jobs came near that target. (Their responses included 899 computer specialists, 452 accountants, and 168 electrical engineers.) All the other series we asked about had fewer respondents. We explored the feasibility of locating larger numbers in the specific needed occupations. Schools provide the Department with lists of graduates to use in sampling. The lists also show each graduate's major field of study. Past surveys can be used to estimate the association between field of study and first job which can be used to guide oversampling in future studies, so that, for example, a certain number of additional math, science, or engineering majors can be sampled to obtain an estimated number of additional respondents who are scientists and engineers.

After we completed our data-gathering, department officials told us they decided to discontinue the specific Recent College Graduates Survey and to establish in 1990 a new longitudinal study involving surveys of students while in school and also 1, 3, and 6 years after graduation. The feasibility and cost of using the new survey as a source of the nonfederal comparison group can be explored when planning for the 1990 survey begins. Specific issues would include the need for questions on educational background and academic achievement, work experience if any, and specific occupation; and the sample would need to yield the needed 200 in each of the selected occupations of interest.

Table III.2 summarizes the cost elements discussed above. The display includes only core professional staff; thus it does not include costs of

---

[illegible] For example, OPM could provide a list of new federal employees in the selected occupations, which could be matched with the student lists provided by colleges during the Education Department survey sampling stage, and those found in this way could be oversampled in an attempt to increase the federal employee group for comparative analysis. The department does some automated and hand checking of these student lists already as part of sampling, for instance to insure over-sampling of Hispanic-surnamed individuals. But as only half the 400 schools involved have automated student records, the full matching process, both by computer and by hand, could be costly.

supervision of the office that is assigned the work or support staff in such an office that would be needed for study tasks and report production. It also does not include cost estimates for either alternative approach we have considered to sampling newly-hired nonfederal employees (reaching them either through employer lists kept at BLS or state offices, or through a planned Department of Education survey of recent graduates sampled from school records). (See table III.2.)

Table III.2: Estimated Costs of the Proposed Assessment of Federal Workforce Quality[a]

Cost Element	Cost
Survey design, development, pre-testing, and analysis of data (maximum of 4.5 staff years at $472 per day)	424,800
Sampling (OPM internal charges)	10,000
Survey administration (64,800 surveys at $1.50)[b]	97,200
Computer costs for analysis (12 months at $3000)	36,000
Estimated contract costs for computer matching on 4,800 new hires at college entrance test firms	50,000
Report production	100,000
Total	**$718,000**

[a] Assumes starting all segments of the assessment and completing one baseline cycle of data gathering and analysis on entering, current, and exiting employees, with comparison groups as discussed in the text. Does not include costs of overall supervision and management, or indirect costs such as space rental, as these could vary widely under different arrangements for conducting the assessment. Also does not include estimated cost of either option for obtaining survey data on comparison group of 2400 newly hired nonfederal employees.

[b] Assumes separated employees can be surveyed at the same cost as current employees, since addresses are kept on file at agencies for several years.

There is an indirect cost to surveys, in the salary paid for the time spent completing the survey, time in which other work is put aside. For the current federal employees proposed to be surveyed in the three parts of the assessment, the total value of the time spent in the surveys is estimated to be $766,187.[2] (In addition, 2,400 people recently hired by nonfederal employers would be surveyed, and also 3,600 former federal employees. We did not attempt to estimate the value of their time spent on the survey.) These costs, which are salary dollars that will be spent anyway, are not shown in table III.2, to avoid confusion with direct dollar costs that would require additional federal expense.

---

[2] The figure assumes that any survey for this assessment would require no more than 45 minutes of duty time to complete. Each sample would contain about half administrative and half professional employees. To calculate the salary cost of the survey, the new-hire sample was assumed to be evenly divided between grades GS-7 and GS-9 (though some unknown fraction of the actual sample would be experienced people hired at higher grades). The overall workforce sample was assumed to be evenly divided between grades GS-11 and GS-12, which are the average grades for administrative and professional occupations. The group of current employees to be matched with those who separated was assumed to be evenly divided between grades GS-12 and GS-14.

Advisory Panel Members, and Others Consulted During Our Work

Advisory Panel

We chose an advisory group for the project consisting of 11 members with expertise in economics, educational and psychological measurement, political science, and evaluation design, as well as experience in Congress, in major business corporations, in federal personnel management and in federal employee groups, at the state level, and in academia. The advisory panel reviewed the project design and approach at the beginning. The group met on January 29, 1988, when we presented tentative results of the work so far and reviewed the basic feasibility of an assessment using samples. Advisory group members completed their work by making comments on a draft of our report. The panel members were:

Dr. Joseph L. Fisher
Special Assistant to the President
George Mason University
Fairfax, VA

Dr. Eli Ginzberg
Conservation of Human Resources
Columbia University
New York, NY

Mr. Jerry Klepner
Director of Legislation
American Federation of State, County and Municipal Employees
Washington, DC

Mr. Paul O'Neill
Chairman of the Board
Aluminum Company of America
Pittsburgh, PA

Professor Nelson Polsby
Department of Political Science
University of California
Berkeley, CA

Mr. J. M. Schulman
Director of Personnel
U.S. Department of Energy
Washington, DC

Professor Lee Sechrest
Department of Psychology
University of Arizona
Tucson, AZ

Dr. Marshall Smith
Dean, School of Education
Stanford University
Stanford, CA

Mr. Ray Sumser
Director of Civilian Personnel
U.S. Army
Washington, DC

Dr. Mary Tenopyr
Selection and Testing Director
AT&T
Short Hills, NJ

Professor Carol Weiss
Graduate School of Education
Harvard University
Cambridge, MA

Consultations With Groups and Individuals

In addition, the development of our approach has benefited from the advice and review of a number of groups and individuals in addition to the advisory panel. We arranged group consultation conferences with three associations in the field:

- American Society for Personnel Administration
- American Society for Public Administration
- International Personnel Management Association, Federal Section

We addressed questions about the issues in assessing the workforce to the heads of four federal employee groups and held individual interviews with staff of each group. These groups were:

- American Federation of Government Employees
- Federal Managers Association
- National Federation of Federal Employees
- National Treasury Employees Union

We met with other groups to present the project and obtain views, including the Panel on Public Service of the National Academy of Public Administration and the Human Resources Task Force of the President's Council on Management Improvement.

In addition, we discussed the problem of workforce quality assessment and various solutions with individuals in several organizations involved with the issues in both public and private sectors. Individuals who made helpful contributions include:

Mark Abramson
Center for Excellence in Government

Sue Berryman
National Center on Education and Employment

Richard Burns
Center for Occupational and Professional Assessment, Educational Testing Service

Dave Crockett
American College Testing Program

Marvin Dunnette
University of Minnesota

Edie Goldenberg
University of Michigan

Glenn Gotz
The Rand Corp., Defense Manpower Research Center

Richard Hackman
Harvard Graduate School of Business

Bernd Hasenkamp
Educational Testing Service

William Kennish
Law School Admissions Council

Steve Kerr
University of Southern California

Allen I. Kraut
IBM

Edward Lawler
University of Southern California

John Lee
National Center for Postsecondary Governance and Finance

Charles Levine
National Commission on the Public Service

Michael Liebman
McManis Associates

Richard Mansfield
McBer and Company, Inc.

Eugene McGregor, Jr.
Indiana University

Brian Morgan
Opinion Research Corporation

James Perry
Indiana University

Karlene Roberts
University of California, Berkeley

Robert Weatherall
Massachusetts Institute of Technology

Alexandra Wigdor
National Research Council

Earlier Correspondence With the Committee

In 1986, the Chair of the House Subcommittee on Civil Service circulated for comment an earlier informal GAO communication on assessing quality in the workforce. Responses from 12 federal agencies as well as many individuals helped shape the plans for the work presented in this report. Some individuals already cited above wrote in response and also subsequently provided additional help; others who sent the committee information and views on ways to assess quality in the workforce were:

Carolyn Ban
State University of New York at Albany

James Bowman
Florida State University

Colin Campbell
Georgetown University

William Carey
American Association for the Advancement of Science

James Fesler
Yale University

Gregory Gaertner
WESTAT Corp.

Michael Hansen
American University

Sar Levitan
George Washington University

Paul Lorentzen
University of Southern California

Bradley Patterson
American Society for Public Administration

Barbara Romzek
University of Kansas

Bernard Rosen
American University

Richard Schmidt
Scanlon, Hastings and Schmidt

David Stanley
Vienna, VA

James Sundquist
Brookings Institution

Frederick Thayer
University of Pittsburgh

Frank Yeager
EDA Systems

(973614)

*U.S. G.P.O. 1988-201-749:80209

Requests for copies of GAO reports should be sent to:

U.S. General Accounting Office
Post Office Box 6015
Gaithersburg, Maryland 20877

Telephone 202-275-6241

The first five copies of each report are free. Additional copies are $2.00 each.

There is a 25% discount on orders for 100 or more copies mailed to a single address.

Orders must be prepaid by cash or by check or money order made out to the Superintendent of Documents.